numero 133
settembre 1979
anno XII
volume XXIII

LE SCIENZE
edizione italiana di SCIENTIFIC AMERICAN

Effetti immediati e ritardati di una guerra nucleare

In caso di una guerra nucleare totale il livello di devastazione dei paesi contendenti sarebbe così catastrofico da rendere priva di senso ogni idea di «vittoria» e da sconvolgere gravemente l'intero pianeta

di Kevin N. Lewis

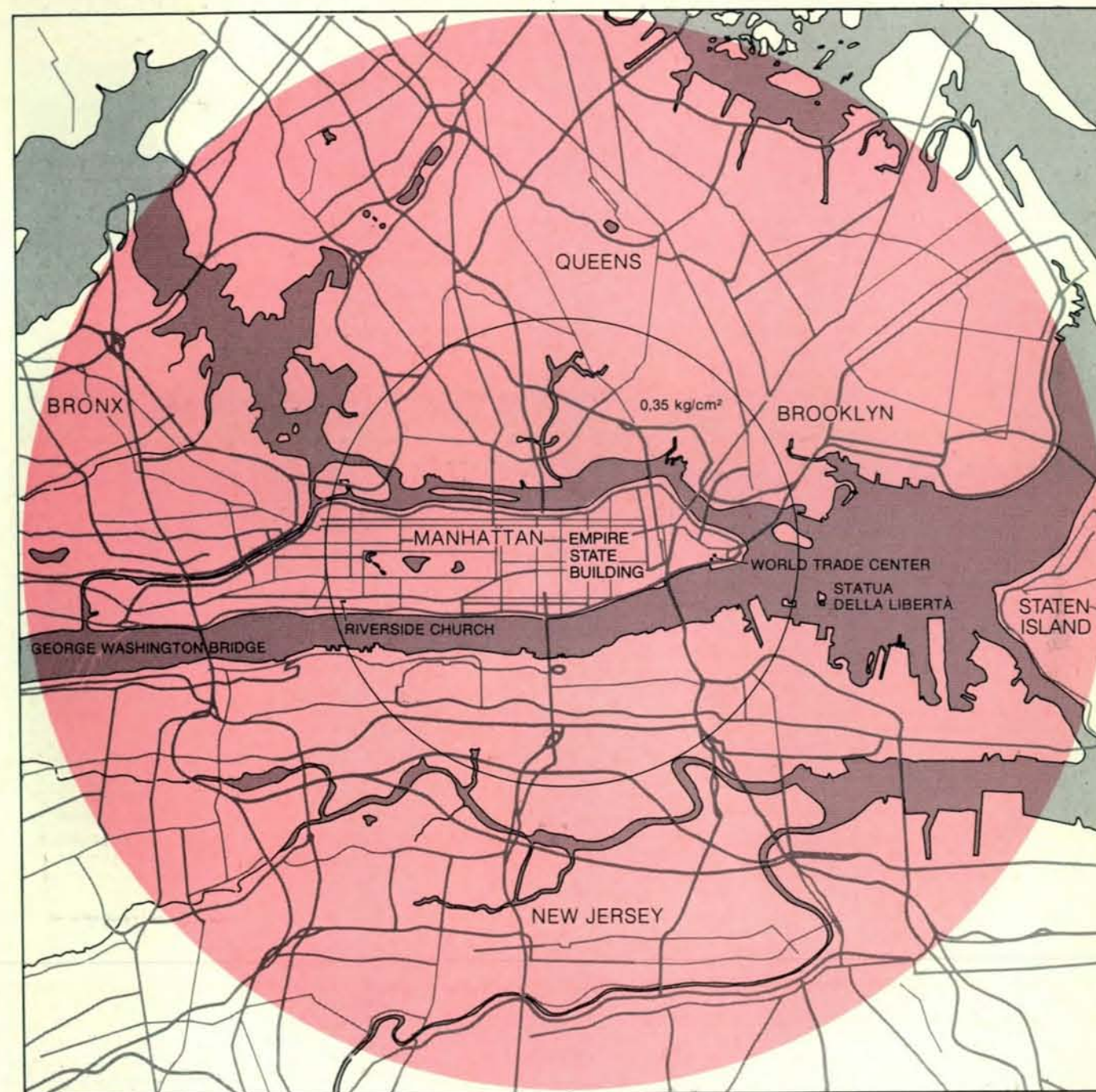
Lo scopo principale delle forze strategiche nucleari di una grande potenza è quello di dissuadere un'altra potenza dallo sferrare un attacco contro di essa o contro i suoi alleati. Per essere all'altezza di questo compito, gli Stati Uniti intendono conservare costantemente la capacità di infliggere all'URSS danni intollerabili. Le forze americane di missili e di bombardieri a lungo raggio sono state ideate pertanto in modo da poter sopravvivere a un attacco generale di sorpresa dell'URSS in numero sufficiente a sferrare un contrattacco distruttivo di rappresaglia. Poiché anche l'URSS dispone di forze analoghe, si ritiene improbabile che una delle due parti trovi vantaggioso attaccare l'altra. È a questo potenziale di ritorsione reciproca, o capacità di distruzione sicura, che si affida in larga misura l'equilibrio militare strategicamente stabile fra le due superpotenze.

Dal momento che, in questa prospettiva, la possibilità di evitare una guerra dipende in parte dal fatto che gli Stati Uniti riescano a mantenere intatta la loro capacità di distruzione sicura, qualsiasi riduzione di tale capacità sarebbe una faccenda seria. Grande preoccupazione hanno suscitato pertanto le recenti dichiarazioni di alcuni analisti militari americani, secondo cui l'URSS sta perseguendo attivamente misure tali da ridurre l'efficacia di un attacco di rappresaglia americano. Si afferma specificamente che ambiziose iniziative russe nel campo della difesa civile creerebbero una pericolosa asimmetria strategica in mancanza di azioni americane che facciano da contrappeso. Per esempio, in concomitanza con un attacco *counterforce* di sorpresa contro i missili americani con base di lancio a terra, l'URSS potrebbe cercare di evacuare le proprie città, col risultato che, in uno scontro nucleare totale, le perdite sovietiche sarebbero notevolmente inferiori a quelle americane. In una situazione del genere, gli Stati Uniti potrebbero trovarsi con le mani legate, non potendo intensificare ulteriormente le ostilità, e l'URSS sarebbe quindi in grado in vari modi di imporre la propria volontà. Anche se dovesse seguirne una guerra totale, l'Unione Sovietica, si dice, sarebbe in grado di riprendersi molto più in fretta degli USA. Queste argomentazioni hanno avuto fra l'altro il risultato di ridestare l'interesse per il moribondo programma americano di difesa civile e di far prendere in considerazione nuove opzioni in merito agli obiettivi dei missili strategici, nell'intento di sconfiggere il programma russo di difesa civile.

Queste sceneggiature ipotetiche si basano in parte su una sottovalutazione dei danni che le forze americane sopravvissute potrebbero infliggere all'URSS. Molte di queste stime tengono conto soltanto degli effetti dirompenti, facilmente calcolabili, delle armi nucleari, senza considerare gli effetti altrettanto distruttivi delle radiazioni termiche e delle radiazioni ionizzanti. Quando nei calcoli si introducono anche questi effetti aggiuntivi, è chiaro che la guerra nucleare rimane un disastro reciproco bell'e buono e che non c'è neanche da pensare a possibili tentativi di difesa civile capaci di modificare sostanzialmente la prospettiva. Pertanto, dal punto di vista militare operativo, le affermazioni secondo le quali la capacità di rappresaglia degli USA si sta «erodendo» sono prive di qualsiasi valore. Inoltre è molto improbabile che la situazione possa cambiare in un futuro prevedibile.

Come vengono stimati i danni di una guerra nucleare, e quali conseguenze di una guerra del genere vengono normalmente escluse da tali stime? L'analisi più completa dimostra che né gli USA né l'URSS hanno motivo di preoccuparsi dell'integrità della propria capacità di rappresaglia. Anche se gran parte dell'attuale dibattito sulla gravità della minaccia sovietica tende a ignorare questo fatto, non può esserci il minimo dubbio che una guerra totale rimanga un progetto perdente per entrambe le parti. Va da sé che un potere di dissuasione credibile si basa su molti fattori oltre alla capacità di sferrare un massiccio attacco di rappresaglia. È per altro nell'interesse di entrambe le parti che dal dibattito strategico venga eliminata l'idea di «vincere» una guerra nucleare totale, nel senso di poter migliorare a un prezzo accettabile la propria posizione relativa, e che tutte le conseguenze di un tale evento rovinoso siano presentate all'attenzione di tutti.

I criteri specifici per valutare l'efficacia di un eventuale attacco di rappresaglia furono stabiliti agli inizi degli anni sessanta sotto la direzione dell'allora segretario alla difesa Robert S. McNamara. Fino a quel momento gli strateghi americani non disponevano di parametri quantitativi formali per determinare i livelli più opportuni delle forze di rappresaglia degli Stati Uniti. McNamara creò allora il concetto di distruzione sicura, sostenendo che la distruzione del 20-25 per cento della popolazione dell'URSS e del 50 per cento almeno della sua capacità industriale avrebbe costituito un danno inaccettabile agli occhi dei capi di quel paese. Stabilendo queste misure, McNamara poté coordinare meglio i programmi della marina e dell'aviazione, per adeguare le esigenze strategiche militari alla struttura esistente delle forze e per eliminare programmi superflui. Sebbene la definizione di un certo livello di danni avesse un intento politico, quello cioè di far pesare sul governo dell'URSS la minaccia di distruzioni intollerabili, la scelta di quelle percentuali specifiche rispecchiava le capacità delle forze strategiche americane programmate a quel tempo. I necessari livelli di distruzione si basavano anche in una certa misura sulle particolari caratteristiche degli obiettivi da colpire nell'Unione Sovietica.

La popolazione e le risorse economiche

Per gli strateghi americani della guerra nucleare, l'area letale è la zona circolare all'interno della quale il numero dei sopravvissuti a un'esplosione nucleare è pari a quello delle vittime che si hanno al di fuori di tale zona. In base a questa premessa semplificante, si può calcolare il numero delle vittime di un'esplosione nucleare moltiplicando l'area letale per la densità della popolazione (ammesso che nell'intera zona la densità sia uniforme). Come regola generale, si fa conto che l'area letale si estenda grosso modo fino al limite della sovrapressione di 0,35 kg/cm² che per l'esplosione in aria da un megaton raffigurata a pagina 10 corrisponde a una superficie circolare di 6,8 chilometri di raggio *(area all'interno del cerchio in nero)*. Il concetto di area letale non include parecchi effetti ritardati importanti (anche se meno prevedibili) delle esplosioni nucleari, come gli incendi e la ricaduta radioattiva. In un giorno sereno, per esempio, un'esplosione in aria da un megaton potrebbe provocare incendi fino a 10 chilometri di distanza. Se questi incendi dovessero consolidarsi in un incendio generale su vasta scala, tutta la zona entro quel raggio *(area in colore)* verrebbe devastata, col risultato di allargare di ben cinque volte l'area letale.

La sequenza della pagina a fronte illustra in ordine cronologico gli effetti immediati dell'esplosione di una testata nucleare da un megaton a circa 2000 metri di altezza sul cuore di New York. Subito dopo l'esplosione, si formerebbe una palla di fuoco estremamente calda e luminosa, la quale creerebbe un'intensa radiazione *(in colore)*, capace di produrre ustioni e di provocare incendi a notevole distanza. L'esplosione darebbe luogo anche a un'onda d'urto distruttiva, che si propagherebbe dalla palla di fuoco a velocità supersonica; 1,8 secondi dopo la detonazione, per esempio, il fronte dell'onda *(cerchio nero)* si troverebbe più o meno a 800 metri dalla palla di fuoco. Inoltre i processi nucleari responsabili dell'esplosione sarebbero accompagnati da un'emissione radioattiva molto intensa, costituita soprattutto da raggi gamma e neutroni *(linee bianche ondulate)*, che nell'aria avrebbe un raggio d'azione sufficiente a farla giungere al suolo nella zona del centro. Quando l'onda d'urto primaria dell'esplosione colpisse il terreno, si formerebbe per riflessione un'altra onda d'urto. A una certa distanza dal punto zero a terra (distanza che varia a seconda dell'altezza dell'esplosione e della potenza della bomba), il fronte dell'onda primaria e quello dell'onda riflessa si fonderebbero, formando un'unica onda di Mach rinforzata; nel caso di una testata da un megaton fatta esplodere a 2000 metri di altezza, l'effetto Mach incomincerebbe circa 4,6 secondi dopo la detonazione a una distanza di 2 chilometri dal punto zero a terra. A quel punto la sovrapressione (vale a dire la pressione dell'aria al di sopra della pressione ambiente atmosferica) sarebbe di 1,12 chilogrammi per centimetro quadrato. 11 secondi dopo la detonazione, l'onda di Mach si sarebbe allontanata dal punto zero a terra di 5,1 chilometri, la sovrapressione all'altezza di tale onda sarebbe di 0,42 kg/cm^2 e la velocità del vento subito dietro l'onda stessa sarebbe di circa 288 chilometri all'ora; quantità cospicue di radiazioni termiche e nucleari continuerebbero ad arrivare al suolo. 37 secondi dopo la detonazione, l'onda di Mach sarebbe quasi a 15,2 chilometri dal punto zero a terra, la sovrapressione all'altezza dell'onda sarebbe di 0,07 kg/cm^2 e la velocità del vento dietro l'onda sarebbe di 64 chilometri all'ora. (I vetri incomincerebbero a rompersi a partire da una sovrapressione dell'ordine di 0,035 kg/cm^2). Anche se la radiazione termica non avesse più valori rilevanti, i raggi gamma continuerebbero a giungere al suolo in quantità potenzialmente letali. La palla di fuoco non sarebbe più luminosa, ma sarebbe sempre molto calda, e quindi prenderebbe rapidamente quota, attirando l'aria verso il proprio interno e verso l'alto e producendo forti correnti d'aria, dette venti secondari, che solleverebbero i detriti e le macerie della città, formando il gambo di quella che alla fine diventerebbe la caratteristica nube a fungo. 110 secondi dopo la detonazione, il residuo caldo della palla di fuoco, pur continuando a salire, incomincerebbe ormai a espandersi e a raffreddarsi. Come risultato, i prodotti vaporizzati provenienti dalla fissione e altri residui della bomba si condenserebbero, formando una nube di particelle radioattive. In questo momento la nube sarebbe arrivata già a 11,2 chilometri di quota. L'altezza massima presumibilmente raggiunta dalla nube (dopo 10 minuti) sarebbe di circa 22,4 chilometri. Alla fine le particelle della nube sarebbero disperse dal vento e, a meno di precipitazioni atmosferiche, non vi sarebbe nessuna ricaduta radioattiva precoce (o locale).

dell'URSS sono concentrate in un numero straordinariamente ridotto di centri urbani di un certo rilievo. Infatti circa un terzo della popolazione e quasi due terzi della capacità industriale del paese sono concentrati nelle 200 città più grandi. Attacchi nucleari contro altre città non accrescerebbero in misura notevole i danni della rappresaglia (eccezion fatta per gli effetti ritardati del *fallout* radioattivo). Il concetto di distruzione sicura introdotta da McNamara potrebbe quindi essere tradotto nella capacità di distruggere le 200 città più grandi dell'URSS.

Data questa premessa, si potrebbero stabilire le esigenze delle forze americane determinando il numero delle testate nucleari necessarie per distruggere gli obiettivi sociali ed economici più importanti in quelle 200 città. La programmazione degli obiettivi da colpire è sensibile a molti fattori operativi, quali la composizione e lo schema planimetrico delle città, ma richiede soprattutto una previsione precisa del modo in cui la popolazione locale sarà colpita dagli effetti letali delle esplosioni nucleari. In pratica, per prevedere i danni di un attacco di rappresaglia si fa un raffronto, città per città, fra le proprietà fisiche delle esplosioni nucleari e le caratteristiche rilevanti dell'obiettivo. I risultati di questi calcoli particolareggiati possono essere per altro approssimativamente abbastanza buoni con l'ausilio di alcune semplici procedure.

La forza esplosiva di un'arma nucleare viene definita in genere in termini di quantità di esplosivo chimico necessaria per liberare un'equivalente quantità di energia; si dice per esempio che una bomba nucleare ha una potenza di un certo numero di chiloton (migliaia di tonnellate) o di megaton (milioni di tonnellate) di tritolo. Come in un'esplosione chimica, l'energia di un'esplosione nucleare viene generata molto rapidamente in una massa ridotta. Quando una bomba nucleare viene fatta esplodere in aria, l'energia liberata istantaneamente vaporizza i componenti della testata, creando una palla di fuoco assai calda che si espande rapidamente. L'esplosione ha due effetti immediati che in un attacco contro una città possono essere distruttivi. Primo, a mano a mano che si espande, la palla di fuoco emette un'onda d'urto attraverso il mezzo circostante. Quest'onda d'urto, che si allontana dal punto dell'esplosione a velocità supersonica, provoca danni agli edifici e alle persone dovuti allo spostamento d'aria. Inoltre la caldissima palla di fuoco irradia energia termica, soprattutto fotoni, nella regione visibile e in quella dell'infrarosso dello spettro elettromagnetico, e questa energia può produrre ustioni e incendiare materiali che non sono protetti da qualche schermatura opaca. Grosso modo metà dell'energia della bomba si converte alla fine in spostamenti d'aria di natura meccanica, e circa un terzo viene liberato sotto forma di radiazione termica. Il resto dell'energia è rappresentato da radiazione nucleare immediata e da radiazione nucleare e termica ritardata, nessuna delle quali viene considerata importante in un programma di distruzione sicura, ma che, tutte insieme, contribuiscono ad aumentare la distruttività di un attacco nucleare.

I movimenti meccanici di un'esplosione nucleare sono analoghi a quelli di un'onda di maremoto. Il fronte d'urto è letteralmente un muro di aria compressa. Al suo passaggio, gli edifici sono esposti a un aumento quasi istantaneo della pressione atmosferica locale e possono sgretolarsi. Dietro il fronte d'urto si muovono venti violenti, analoghi alle correnti d'acqua che seguono il moto di un'onda dell'oceano. Le forze prodotte da questi venti possono condurre anche al crollo di edifici nella zona in cui si trovano gli obiettivi da colpire. A seconda della loro forma e della loro costruzione, gli edifici possono essere vulnerabili o all'onda d'urto o ai venti che la seguono o all'una e agli altri. La «durezza» di un obiettivo (vale a dire la sua capacità di resistere all'effetto distruttivo dell'onda d'urto) viene definita in genere dalla «sovrapressione» indotta (espressa in chilogrammi per centimetro quadrato al di sopra della pressione atmosferica) a cui l'obiettivo viene distrutto.

La radiazione termica può fare vittime direttamente per ustioni dovute al calore o, indirettamente (dando fuoco ai materiali vicini), vittime per ustioni dovute alle fiamme, e queste vittime si aggiungono a quelle colpite dall'effetto dirompente. L'entità di questi danni dipende sia dalla potenza dell'energia radiante liberata (misurata in genere in calorie per centimetro quadrato), sia dalla durata di emissione dell'energia stessa. Gli effetti distruttivi dello spostamento d'aria diminuiscono con la distanza più rapidamente degli effetti termici. In condizioni ideali, quindi, un'esplosione nucleare può provocare notevoli danni incendiari ben oltre la zona devastata dagli effetti dirompenti. I danni termici per altro sono molto soggetti all'influsso di fattori esterni, fra cui la presenza di nubi o di un manto di neve, la trasparenza relativa dell'atmosfera e le caratteristiche del bersaglio. Perciò gli effetti termici sono molto meno prevedibili degli effetti dirompenti diretti.

Poiché le forze di rappresaglia sono progettate in modo da poter infliggere danni sicuri, le conseguenze di un attacco sono calcolate tipicamente solo sulla base dei più prevedibili effetti dirompenti. Si consideri il problema di distribuire un «pacchetto» opportuno di bombe o testate nucleari su un'area urbana dopo un attento esame degli obiettivi da colpire nella città. I punti di mira di ogni bomba o testata vengono scelti in modo tale da assicurare che gli effetti desiderati dello spostamento d'aria raggiungano tutti gli obiettivi. Se questi sono abbastanza vicini fra loro, potrebbe bastare una sola testata. Se invece sono dispersi o «duri», più che aumentare la carica esplosiva di una singola bomba, potrebbe forse essere preferibile usare più di una bomba o testata per colpire tutta la zona. Questo metodo si premunisce dal possibile fallimento di un'unica grande testata, che lascerebbe «scoperto» un bersaglio, e rispecchia anche il fatto che pochi complessi industriali e militari sono sufficientemente

concentrati o hanno la forma giusta per essere attaccati con una sola arma del tipo che attualmente costituisce il grosso dell'arsenale strategico americano.

Ogni città ha una serie unica nel suo genere di obiettivi con determinate caratteristiche, ma alcune semplici norme rendono possibile prevedere l'ammontare dei danni e delle vittime. In linea generale, qualsiasi fabbricato che non fosse costruito specificamente per resistere all'effetto dirompente verrebbe distrutto se fosse soggetto a una sovrapressione di 0,35 o più chilogrammi per centimetro quadrato oltre la pressione atmosferica ambientale, che è di circa un chilogrammo per centimetro quadrato, e quelle costruzioni che non crollassero subirebbero comunque danni tali da impedire qualsiasi lavoro di riparazione. Alcuni edifici con struttura portante in cemento armato (e materiali pesanti all'interno) potrebbero resistere a una sovrapressione di 2,8 chilogrammi o più per centimetro quadrato, ma se questi obiettivi fossero ritenuti importanti, l'attaccante potrebbe abbassare la quota a cui far esplodere la sua bomba o indirizzare le sue armi d'offesa (o aumentare il numero delle testate) in modo da raggiungere gli effetti desiderati. A occhio e croce tuttavia una sovrapressione di 0,35 chilogrammi per centimetro quadrato è considerata sufficiente a distruggere la maggior parte degli edifici.

Il corpo umano può sopportare spostamenti d'aria molto più intensi di quasi tutti i fabbricati, per cui in un attacco nucleare la maggior parte delle vittime dell'effetto dirompente sarebbero causate in maniera indiretta. Il grosso della popolazione correrebbe il pericolo di trovarsi all'interno o nelle vicinanze di edifici che crollano, di essere colpita dai detriti lanciati dall'onda d'urto o di essere scagliata contro una superficie immobile. Anche gli effetti termici causerebbero molte vittime entro un certo raggio, indipendentemente dalle condizioni esterne. Nel calcolare in anticipo il numero delle vittime, si fa uso spesso del concetto semplificante di «area letale». Basata su dati teorici ed empirici elaborati dall'Atomic Energy Commission agli inizi degli anni cinquanta, l'area letale è la zona circolare entro la quale il numero dei sopravvissuti sarebbe pari a quello delle vittime fuori del cerchio, ammesso che la densità della popolazione nell'intera area sia uniforme.

Per semplificare i calcoli, le perdite stimate vengono ridistribuite, per cui gli strateghi calcolano fra le vittime tutti coloro che si trovano all'interno del cerchio, escludendo quanti ne sono fuori. Una stima delle vittime immediate si ha quindi moltiplicando l'area letale per la densità della popolazione. L'esperienza di Hiroshima e di Nagasaki e i dati degli esperimenti nucleari indicano che per una bomba dell'ordine di 20 chiloton l'area letale si estende grosso modo alla zona in cui vi è una sovrapressione di un chilogrammo per centimetro quadrato. Pertanto la copertura offerta da tale sovrapressione è considerata un parametro soddisfacente per calcolare sia il numero delle vittime, sia la distruttività economica delle esplosioni nucleari.

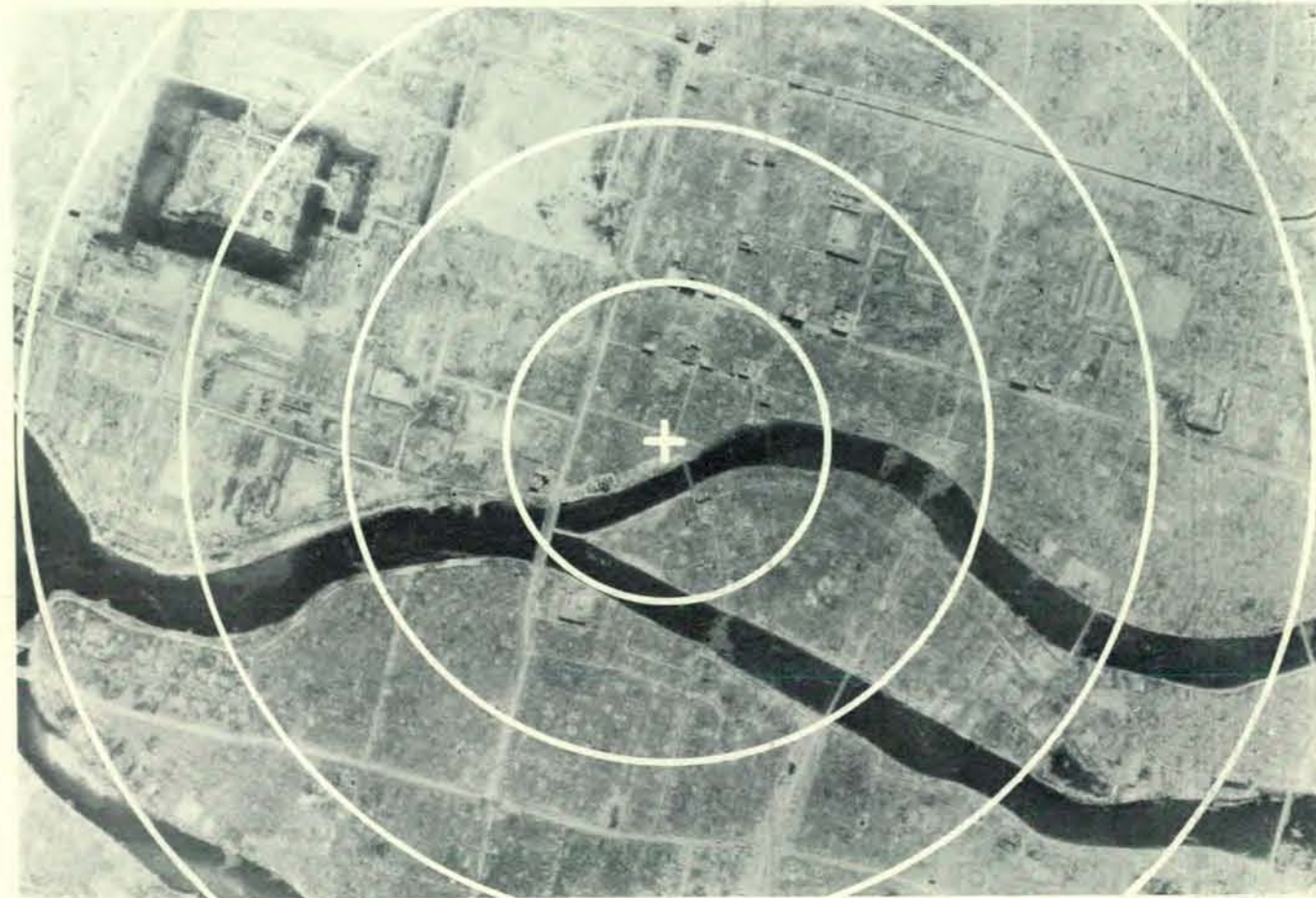
Le bombe o le testate nucleari generano una sovrapressione di 0,35 kg/cm^2 a una distanza proporzionale alla radice cubica della loro forza esplosiva. Per questa ragione si dice che le bombe più grandi distribuiscono la loro potenza distruttiva in maniera molto meno efficace di quelle più piccole. Una bomba da 100 chiloton, per esempio, genera una sovrapressione di 0,35 kg/cm^2 entro un raggio di circa 3200 metri. Ebbene, una testata con una forza esplosiva dieci volte superiore (un megaton) genera la stessa sovrapressione soltanto al doppio di tale distanza. In riconoscimento dell'efficacia intrinsecamente maggiore delle bombe più piccole, una misura graduata nota col nome di «megaton equivalenti», definita come la forza esplosiva di una bomba in megaton elevata alla potenza di due terzi, è considerata un indice di capacità distruttiva contro una città migliore della semplice forza esplosiva espressa in megaton. Negli anni sessanta gli specialisti di analisi dei sistemi di McNamara hanno calcolato che 400 megaton equivalenti ucciderebbero il 30 per cento della popolazione e distruggerebbero il 75 per cento della capacità industriale dell'URSS; di recente è stato calcolato che i danni alla popolazione e alle industrie siano più vicini rispettivamente al 35 e al 70 per cento.

Gli effetti immediati e ritardati delle bombe o testate nucleari si possono mettere meglio in rilievo sullo sfondo di un ipotetico attacco a un tipico obiettivo urbano, l'area metropolitana della grande Boston. L'esplosione di 10 testate da un megaton, destinate a obiettivi economici e militari del luogo, genererebbe una sovrapressione di 0,35 kg/cm^2 su una superficie di oltre 1300 chilometri quadrati. Più di 1,3 milioni di persone verrebbero uccise dall'effetto dirompente e dagli effetti termici immediati delle esplosioni, e più dell'80 per cento della capacità industriale della zona andrebbe distrutta. È probabile che gli effetti secondari delle esplosioni, con particolare riferimento agli incendi e alla ricaduta radioattiva, aumenterebbero questi totali.

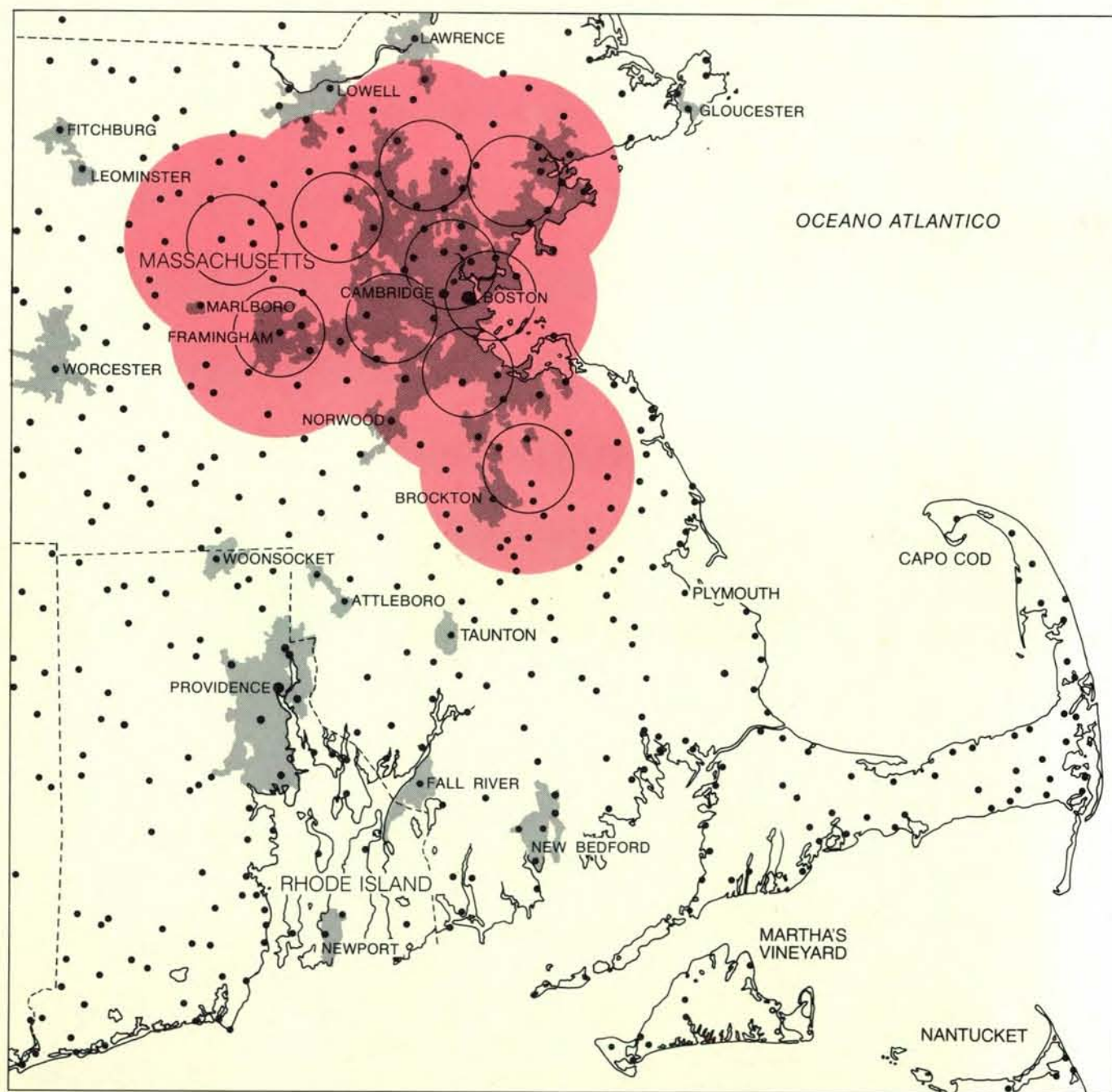
Se le condizioni fossero favorevoli all'attacco, l'effetto più distruttivo potrebbe essere quello di natura incendiaria. In certe condizioni di tempo, ogni esplosione da un megaton provocherebbe incendi fino a 16 chilometri di distanza. In un attacco del genere la minaccia di incendi incomberebbe presumibilmente su buona parte del Massachusetts orientale. Agli incendi provocati dalla violenta emissione di calore si unirebbero quelli dovuti a stufe, caldaie e forni rovesciati dallo spostamento d'aria. La dispersione dei detriti e la rottura di condutture e di serbatoi aggiungerebbero combustibile al fuoco. Gli incendi si propagherebbero a causa del materiale scagliato a distanza dallo spostamento d'aria. Dopo l'attacco, l'estinzione di centinaia di piccoli incendi sarebbe un compito immane. Le condutture dell'acqua salterebbero, le attrezzature anti-incendio andrebbero distrutte e i vigili del fuoco sarebbero messi nell'impossibilità di agire. A Hiroshima il 70 per cento circa delle attrezzature anti-incendio andò distrutto nel crollo delle caserme dei pompieri e l'80 per cento dei vigili del fuoco non si presentò all'appello.

A seconda delle condizioni del tempo e delle caratteristiche della zona in cui si trovano i bersagli da colpire (con particolare riferimento alla densità delle strutture infiammabili), i molti incendi singoli potrebbero consolidarsi fino a formare uno di questi due tipi di incendi a grande scala: la tempesta di fuoco o la conflagrazione. Una tempesta di fuoco è attivata da una forte corrente ascensionale di aria calda, che viene sostituita da aria fredda aspirata dalla periferia dell'incendio. Una conflagrazione è attivata anche da un forte vento di terra già presente prima dell'attacco. Mentre una tempesta di fuoco continua soltanto finché continuano i suoi venti centripeti, una conflagrazione può continuare finché c'è combustibile.

La conseguenza di un incendio a grande scala è la totale devastazione dell'area colpita. In un incendio di questo genere la temperatura può superare i 1000 gradi centigradi, una temperatura più alta di quella necessaria per fondere il vetro e il metallo e per bruciare materiali normalmente indistruttibili. A Hiroshima una bomba atomica con una forza esplosiva dell'ordine di 15 chiloton provocò una tempesta di fuoco che durò sei ore, distruggendo completamente la città per 12 chilometri quadrati. Le città americane sono costruite con materiali che resistono al fuoco più di quelli di Hiroshima, ma per contro hanno un maggior numero di edifici e una maggiore quantità di combustibili, soprattutto benzina e gasolio per riscaldamento, pronti ad alimentare gli incendi. Ancora più importante è il fatto che la forza esplosiva di molte armi nucleari strategiche moderne supera di due

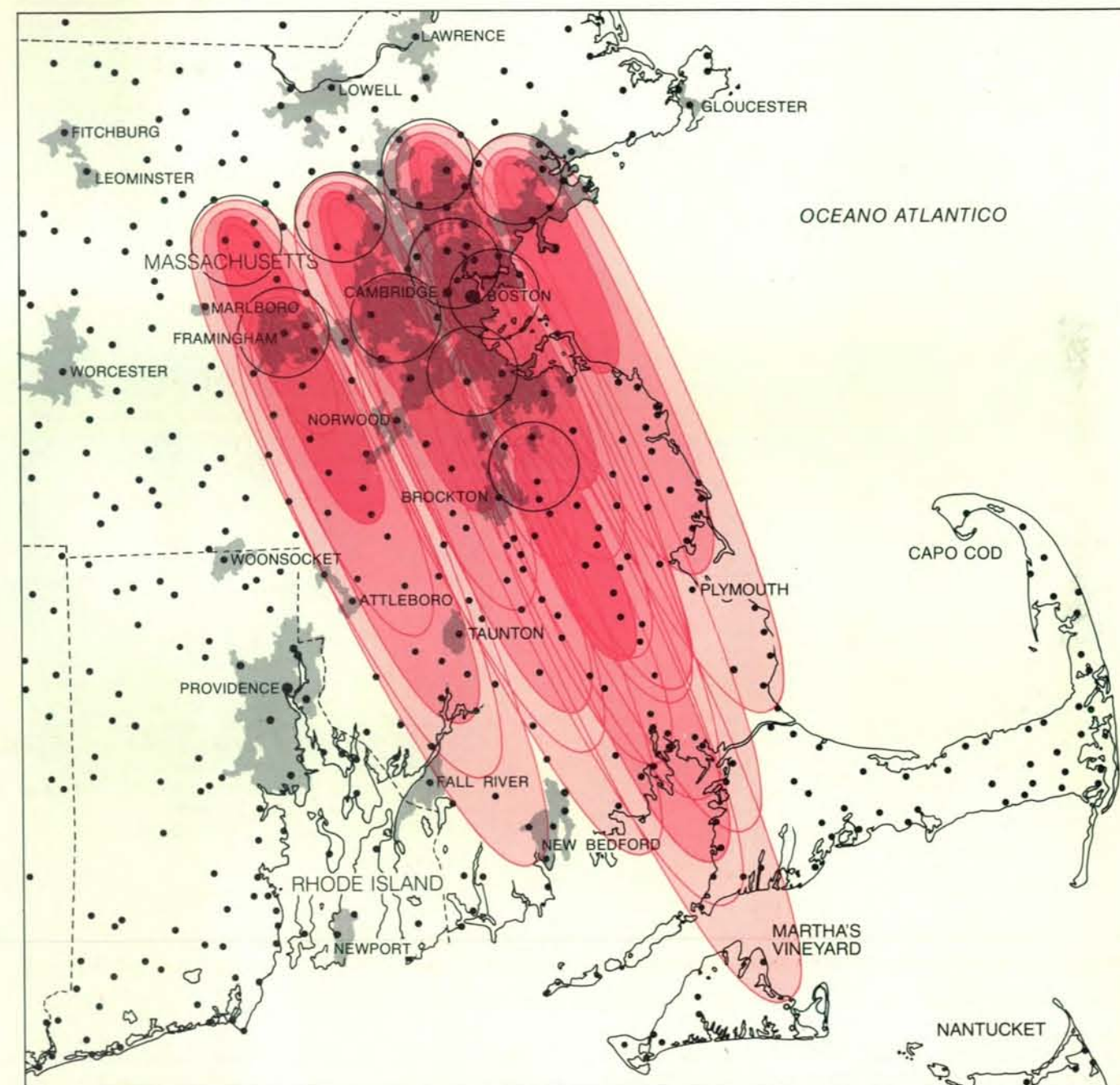
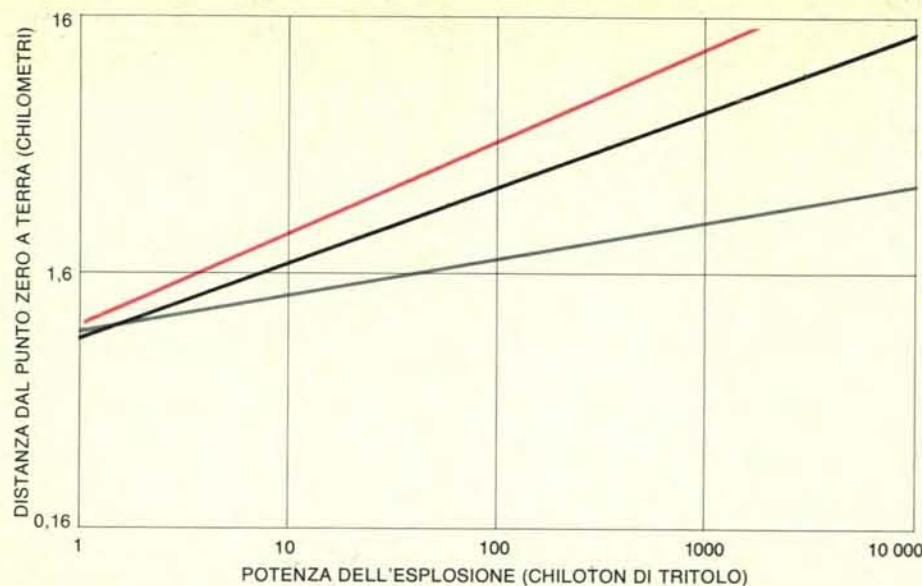
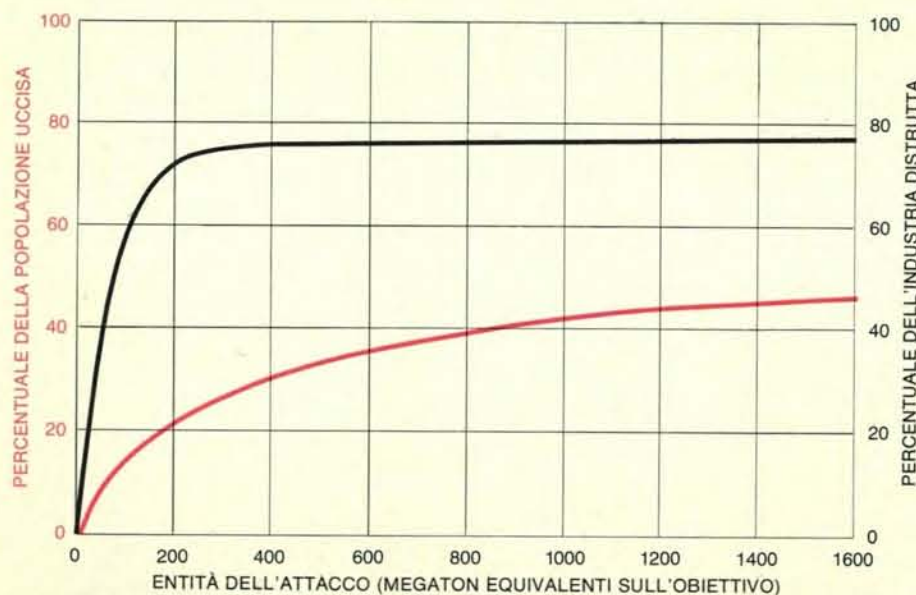
Hiroshima vista dall'alto in queste fotografie prese da un aereo da ricognizione americano prima *(a sinistra)* e dopo *(a destra)* il bombardamento atomico della città, avvenuto il 6 agosto 1945. La croce indica il punto zero a terra, il punto sul terreno proprio sotto l'esplosione.

I cerchi concentrici segnano intervalli di 304 metri. La tempesta di fuoco che seguì gli effetti immediati dell'esplosione durò circa 6 ore e distrusse 11,17 chilometri quadrati della città. La forza esplosiva della bomba che provocò questa devastazione era dell'ordine di 15 chiloton.

o più ordini di grandezza quella delle bombe esplose a Hiroshima e Nagasaki. Inoltre una buona parte della zona sottoposta all'attacco sarebbe esposta alle radiazioni termiche di più palle di fuoco.

I rifugi intesi a proteggere dagli effetti dirompenti non servirebbero molto contro i grandi incendi. La sopravvivenza di chi vi si ripara dipenderebbe in maniera critica dalla temperatura e dall'umidità all'interno del rifugio e, nel caso in cui dovessero scoppiare incendi su larga scala, il problema di mantenere nel rifugio stesso un ambiente in cui la gente possa sopravvivere si aggraverebbe fino al punto di non presentare nessuna soluzione. Inoltre, a meno che non vi fosse una provvista indipendente di ossigeno per ogni singolo rifugio, l'anidride carbonica e altri gas tossici generati dall'incendio potrebbero rivelarsi mortali per coloro che si trovano dentro. Il riscaldamento dei rifugi, sia a opera delle fiamme sia a opera delle macerie riscaldate (che potrebbero rimanere intollerabilmente calde per giorni e giorni dopo la fine dell'incendio), metterebbe in pericolo la vita di chi si trova in un rifugio con un'atmosfera isolata. A Dresda, dove una tempesta di fuoco provocata da bombe di tipo tradizionale uccise nel 1945 più di 100 000 persone, soltanto gli abitanti che avevano lasciato i loro rifugi prima che incominciasse la tempesta di fuoco riuscirono a sopravvivere alla duplice minaccia dei gas nocivi e del riscaldamento dei rifugi.

In questo attacco ipotetico contro l'area metropolitana della grande Boston sono contrapposti gli effetti immediati e gli effetti ritardati di molteplici esplosioni nucleari. In entrambi i casi l'attacco è costituito dalla detonazione di 10 testate nucleari da un megaton dirette contro obiettivi economici e militari. Nella figura a sinistra si è partiti dal presupposto che tutte le bombe vengano fatte esplodere a una quota scelta in modo da massimizzare gli effetti termici e quelli prodotti dall'effetto dirompente. I contorni circolari in nero corrispondono alle zone esposte a una sovrapressione di almeno 0,35 kg/cm²; ognuna di queste zone ha un raggio di 6,8 chilometri. Le zone in colore sono quelle esposte a gravi rischi di incendi e di ustioni in una giornata serena; ogni area in questo caso ha un raggio di 16 chilometri. Qui il principale effetto ritardato è il rischio di una tempesta di fuoco o di una conflagrazione, che interesserebbero tutta la zona e che potrebbero accrescere di 500 000 unità il numero delle vittime degli effetti termici e di quelli prodotti dall'effetto dirompente delle esplosioni nucleari,

Dopo un attacco nucleare molta gente rimarrebbe mutilata, intrappolata fra le macerie o impossibilitata a fuggire dalla città per via delle strade bloccate dai detriti e dal fuoco. Se si formassero incendi su grande scala, eventualità probabile nel caso di molte esplosioni da un megaton, pochi potrebbero essere i sopravvissuti fra coloro che riuscissero a sfuggire a una mutilazione immediata. Se, per esempio, si dovessero verificare incendi su larga scala nell'area di Boston, il numero delle vittime potrebbe aumentare di 500 000 unità.

Un altro fattore che non è incluso nei calcoli relativi alla distruzione sicura è il *fallout* radioattivo. Questa ricaduta radioattiva è dovuta alla condensazione dei sottoprodotti radioattivi di un'esplosione nucleare sui materiali fusi dal calore intenso della palla di fuoco e (in misura molto minore) dalla conversione di materiali non radioattivi in materiali radioattivi a causa dell'assorbimento di neutroni provenienti dalle reazioni nucleari dell'esplosione. Se una bomba nucleare venisse fatta esplodere a livello del suolo o quasi, il *fallout* costituirebbe una grossa minaccia. Grandi quantità di detriti verrebbero sollevate nella nube ascendente, per ricadere più tardi (o più probabilmente essere eliminate dalla pioggia) dalla nube stessa in quantità letali per centinaia di chilometri sottovento. Una dose di radiazione ionizzante dell'ordine di 400-500 rem (il rem è un'unità di misura degli effetti bio-

stimato in 1,3 milioni sulla base del concetto di distruzione sicura. Nella figura a destra il presupposto è che tutte le 10 testate vengano fatte esplodere a livello del suolo in modo da massimizzare gli effetti della ricaduta radioattiva. (I contorni sono stati tracciati sulla base dello schema dei venti tipici predominanti di gennaio.) Le aree in colore intenso sono quelle coperte da una quantità di radiazione che sarebbe fatale almeno all'80 per cento della popolazione esposta. Le aree in colore medio sono quelle in cui almeno il 50 per cento della popolazione esposta morirebbe di qualche malattia da radioattività. Le aree in colore chiaro sono la probabile estensione della zona in cui una buona parte della popolazione esposta presenterebbe sintomi clinici da radioattività, con circa il 20 per cento di casi letali. (I sopravvissuti potrebbero subire gli effetti del fallout degli attacchi in zone limitrofe.) Il numero delle vittime di un attacco condotto con un'opportuna combinazione di bombe nucleari fatte esplodere in aria e al suolo potrebbe essere di oltre due milioni di morti e altrettanti feriti e ammalati.

Il raggio tipico a cui si estendono tre effetti nocivi delle armi nucleari è raffigurato qui per una tipica esplosione in aria in funzione della sua forza esplosiva. La curva in colore rappresenta la distanza alla quale la radiazione termica può causare ustioni di secondo grado alla pelle e provocare incendi, creando il rischio di un incendio generale su vasta scala. La curva in nero misura il raggio del cerchio della sovrapressione di 0,35 kg/cm², all'interno del quale il passaggio del fronte dell'onda d'urto, seguita da venti che soffiano alla velocità di circa 260 chilometri all'ora, provocherebbe una massiccia distruzione della città e un'alta percentuale di vittime. La curva in grigio indica il raggio entro il quale la radiazione nucleare improvvisa dell'esplosione provocherebbe il 100 per cento di vittime. È evidente che, in condizioni di tempo favorevoli, gli effetti termici distruttivi di un'esplosione di questo genere potrebbero andare molto al di là dell'area di massima distruzione dell'effetto dirompente. La radiazione nucleare immediata, d'altro canto, non è chiaramente un importante meccanismo di distruzione per quel che riguarda le armi strategiche nucleari (che hanno una forza esplosiva variabile da qualche decina di chiloton a molti megaton), in quanto l'area coperta dalla radiazione letale sarebbe esposta anche a gravi effetti termici e spostamenti d'aria. Soltanto quando la forza esplosiva è molto più ridotta (dell'ordine di un chiloton o meno) la radiazione nucleare immediata diventa un importante meccanismo letale; questo rapporto è il principio di fondo della bomba a radiazione intensificata, o bomba al neutrone.

Il concetto di distruzione sicura, su cui gli strateghi americani si basano per determinare il potenziale di rappresaglia necessario alle forze nucleari degli Stati Uniti per evitare un attacco di sorpresa dell'URSS, è calibrato qui in termini del numero di megaton equivalenti che ci vorrebbe per distruggere i più importanti centri abitati e gli obiettivi industriali dell'Unione Sovietica. (I megaton equivalenti sono la forza esplosiva di una bomba nucleare elevata alla potenza di 2/3.) Dato il valore decrescente dell'aggiunta di altri megaton equivalenti a un attacco di rappresaglia di questo tipo, è evidente da queste curve che l'invio sull'obiettivo di circa 400 megaton equivalenti sarebbe più che sufficiente per ottenere la distruzione sicura. I danni alla popolazione *(curva in colore)* sono stati stimati soltanto in termini di morti; i danni industriali *(curva in nero)* sono stati determinati calcolando il «valore aggiunto di fabbricazione» distrutto. (Il valore aggiunto di fabbricazione è il valore incrementale conferito alle materie prime in ogni processo industriale.)

logici delle radiazioni sull'uomo) emessa per un periodo di parecchi giorni ucciderebbe metà delle persone che l'avessero ricevuta. Una dose compresa fra 200 e 300 rem ne ucciderebbe poco meno del 20 per cento (ammessa una cura medica immediata), ma gravi alterazioni del sangue connesse alle radiazioni, fra cui una diminuita risposta immunologica, potrebbero accrescere il numero delle vittime aumentando le infezioni letali. Se dieci bombe da un megaton dovessero esplodere a livello del suolo (per massimizzare gli effetti della ricaduta radioattiva più degli effetti termici o di quelli dirompenti), fino a un milione di abitanti del New England che non fossero soggetti agli effetti termici e dirompenti immediati delle esplosioni nucleari sarebbero esposti a pericolosi livelli di radiazioni. Anche con valutazioni ottimistiche sulla disponibilità di rifugi e di viveri, le vittime della ricaduta radioattiva che bisognerebbe aggiungere al totale relativo all'area di Boston potrebbero arrivare a 500 000. Un attacco di questo tipo potrebbe benissimo mescolare le esplosioni in aria e al suolo in modo da creare livelli massimi di entrambi i tipi di danni.

Dopo un attacco di questo genere, il numero delle vittime degli incendi e delle radiazioni continuerebbe ad aumentare, in parte anche perché il personale e i servizi medici andrebbero distrutti. Le vittime di ustioni presenterebbero un problema medico eccezionale, in quanto i casi di ustioni gravi esigono una cura immediata e intensiva, se si vuole che chi ne è colpito sopravviva. La capacità di qualsiasi servizio medico di trattare un numero elevato di vittime di questo genere è limitata anche in tempo di pace. L'afflusso di una cinquantina di persone sopravvissute allo scontro di due reattori di linea nel cielo di Tenerife (isole Canarie) tese al massimo alcuni anni or sono i centri per ustionati degli USA, che hanno una capacità massima di circa 130 pazienti. È ovvio che dopo un attacco nucleare il numero dei casi di ustioni sarebbe infinitamente superiore e l'accesso alle cure mediche sarebbe quindi più difficile.

Sui servizi medici esistenti verrebbe a gravare anche l'incidenza delle ferite e delle lesioni riportate molto al di là delle zone di diffusa mortalità. Il pericolo di ferite provocate da oggetti scagliati violentemente (soprattutto frammenti di finestre in frantumi) esisterebbe a più di 12 chilometri dal centro di uno scoppio da un megaton, e potrebbero aversi ustioni gravi fino a oltre 14 chilometri di distanza, a seconda delle condizioni del tempo. Molte vittime di ustioni, di malattie da radioattività e di altre ferite mortali che non morissero immediatamente avrebbero bisogno di cure mediche intensive (ma non disponibili nella circostanza). Il trattamento dei feriti meno gravi e dei giovanissimi, delle persone molto anziane e di quelle che richiedono cure mediche particolari verrebbe complicato dalla scarsità di cibo, di ricoveri e di medicinali.

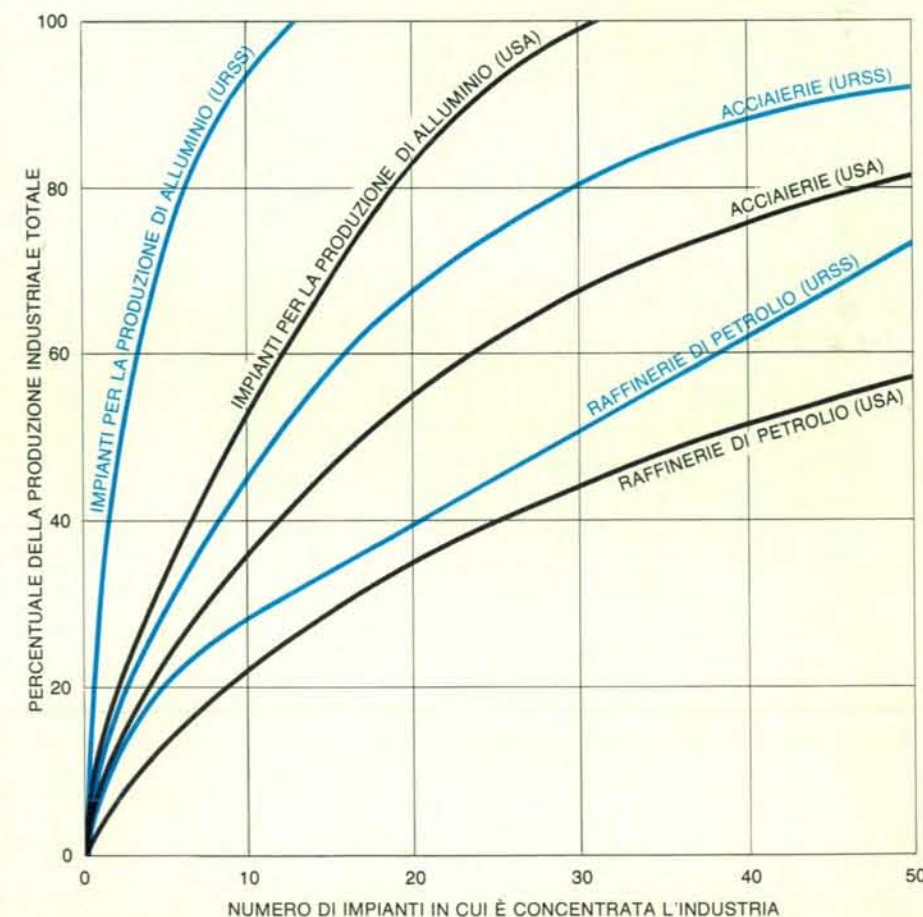
Fra i sopravvissuti a un attacco nucleare totale ve ne sarebbero molti resi permanentemente invalidi da mutilazioni, cecità e altre cause. Qualunque sforzo medico verrebbe ostacolato dalla distruzione del personale e dei servizi sanitari pubblici, dalla proliferazione degli organismi patogeni (che tendono a sopravvivere a elevati livelli di radiazione) e da altre difficoltà, come il problema apparentemente insolubile dell'eliminazione dei morti. Il totale delle vittime di un attacco contro la zona di Boston con bombe nucleari fatte esplodere in aria e al suolo potrebbe benissimo superare i due milioni di morti e un numero grosso modo analogo di feriti e di ammalati.

Il concetto di distruzione sicura trascura anche alcuni problemi di natura strategica. Si dice per esempio che l'URSS stia perseguendo due programmi che le permetterebbero di attenuare l'efficacia di un attacco di rappresaglia degli USA. Il primo programma mira a ridurre il numero delle testate americane dirette contro i loro obiettivi, mediante la distruzione delle forze strategiche USA con un attacco di sorpresa e l'intercettazione del maggior numero possibile di testate prima che queste giungano a bersaglio nell'Unione Sovietica. Il secondo programma tende a minimizzare i danni provocati dalle armi USA in arrivo, mediante l'evacuazione degli abitanti dalle città e la dispersione e l'«indurimento» dei siti industriali. Poiché i sovietici potrebbero quindi «negare» a un attacco americano importanti obiettivi economici e urbani, vi è chi sostiene che le forze americane non assicurerebbero i livelli di danneggiamento previsti dal concetto di distruzione sicura e che l'Unione Sovietica manterrebbe intatti la base industriale, il personale e l'amministrazione necessari per una rapida ripresa postbellica.

Anche ammettendo il pieno successo di questi tentativi sovietici nessuna delle due strategie potrebbe ridurre efficacemente la devastazione di una guerra nucleare totale o accrescere in misura notevole il potenziale di ripresa dell'URSS. Al contrario, questi progetti sembrano soltanto ridurre le capacità di rappresaglia degli Stati Uniti nella prospettiva dell'angusta e arbitraria definizione di distruzione sicura di cui abbiamo discusso in precedenza.

Le analisi di questo genere ignorano il fatto che anche nelle peggiori circostanze gli USA sarebbero in grado di organizzare un attacco di rappresaglia più che adeguato. Qualsiasi piano sovietico tendente a ridurre la capacità americana di distruzione sicura si troverebbe di fronte al compito immane di far scendere le forze americane molto al di sotto dei 400 megaton equivalenti pronti a essere lanciati su un bersaglio. (In raltà, il compito potrebbe essere ancora più difficile, perché un attacco americano ben congegnato di soli 200 megaton equivalenti potrebbe pur sempre uccidere immediatamente un quinto della popolazione e distruggere più di due terzi dell'industria dell'URSS, soddisfacendo con ciò i requisiti della distruzione sicura.)

È estremamente improbabile che un primo colpo preventivo da parte dell'URSS possa raggiungere questa meta. Innanzi tutto, 400 megaton equivalenti rappresentano soltanto una frazione dell'attuale arsenale nucleare americano, costituito da oltre 6000 megaton equivalenti. Più di metà di questo arsenale è pronta in sommergibili lanciamissili all'ancora nelle loro basi, in bombardieri in stato di preallarme nelle basi dello Strategic Air Command e in missili nelle loro basi a terra, tutti in grado di entrare in azione nello spazio di pochi minuti a un ordine del Presidente. Il resto delle forze strategiche americane è costituito soprattutto da bombardieri non in stato di preallarme e da sommergibili in porto per manutenzione. Se un attacco russo di sorpresa dovesse distruggere nei loro silos molti missili americani con base di lancio a terra e tutti i bombardieri e i sommergibili non in stato di preallarme, rimarrebbero sempre a disposizione per un'azione di rappresaglia più di 2000 megaton equivalenti. Anche se un numero inaspettatamente grande di armi USA dovesse funzionare male o essere distrutto in volo, potrebbero pur sempre essere lanciati contro un obiettivo con la massima sicurezza più di 1500 megaton equivalenti. Queste cifre presuppongono «il peggiore dei casi» dal punto di vista degli Stati Uniti: se prima di un attacco sovietico di questo genere vi fosse poi un minimo di preavviso, sarebbe possibile dare l'allarme ad altri bombardieri e sommergibili, per cui il numero dei megaton equivalenti in grado di raggiungere un bersaglio aumenterebbe di oltre il doppio.

Quanto alla disponibilità di quella che a volte è indicata come «sovracapacità distruttiva», un'espressione che intende sottolineare il fatto che le sole bombe nucleari già costruite basterebbero a uccidere più volte ogni essere umano, qualsiasi tentativo di ridurre il numero o l'efficacia delle testate americane in arrivo è destinato a fallire. Mosca, per esempio, è protetta da un sistema di antimissili balistici (ABM) che è limitato per trattato a 100 vettori. (Attualmente solo 64 missili sono impiegati in tale sistema.) Nel caso di un attacco missilistico, quei missili potrebbero distruggere una certa frazione dei missili in arrivo. Gli strateghi americani potrebbero facilmente controbilanciare questo potenziale logoramento con parecchie strategie, una delle quali sarebbe quella di assegnare altre testate al «pac-

Gli obiettivi industriali di vitale importanza risultano più concentrati nell'URSS che negli USA, come risulta da queste tre coppie di curve. Di conseguenza basterebbe un minor numero di testate per paralizzare la produzione sovietica di materiali importantissimi, come acciaio, petrolio e metalli non ferrosi. Inoltre le due economie sono caratterizzate da strozzature cruciali. Per esempio, una sola fabbrica di Pavlodar, nell'URSS, svolge un lavoro essenziale al 65 per cento dell'industria dell'alluminio. Analogamente, quasi l'80 per cento del minerale di ferro imbarcato negli USA è convogliato attraverso le rapide di Sault Sainte Marie, superate con canali navigabili.

chetto Mosca», che si basa su generose premesse teoriche sull'efficacia del sistema ABM moscovita.

Il principio di distruzione sicura assume implicitamente quello che potrebbe essere il quadro meno probabile di una guerra nucleare generale. È estremamente improbabile che una guerra totale incominci con un massiccio attacco di sorpresa di una delle due parti alle città nemiche. È più probabile che una guerra del genere segua una crisi strisciante che potrebbe incominciare con attacchi nucleari contro obiettivi militari. Queste sceneggiature alternative implicano che la popolazione delle città abbia un ampio preavviso di un possibile o probabile attacco nucleare, per cui sarebbe possibile ricorrere all'evacuazione o ad altre tattiche per ridurre i danni. Se la situazione dovesse presentarsi in questi termini, è chiaro che al momento dell'attacco il carattere della popolazione sarebbe manifestamente mutato. Poiché la distruzione sicura viene calcolata in base alla densità della popolazione in tempo di pace e poiché i calcoli si basano su certe premesse relative alla disposizione degli abitanti e dei lavoratori di una città di giorno in giorno, può darsi che in circostanze realistiche non si riesca a raggiungere il progettato numero di vittime.

Il principio di distruzione sicura presuppone anche che una guerra nucleare generale sia costituita da un unico massiccio attacco *«countervalue»*, un attacco cioè diretto contro obiettivi sia militari sia economici. Attacchi countervalue potrebbero per altro mantenersi per un certo tempo a un livello di intensità relativamente basso. Un limitato scambio di colpi countervalue potrebbe consistere in attacchi a località industriali lontane dalle grandi città per scoraggiare *l'escalation* agli attacchi col maggior numero possibile di vittime. (I sovietici in particolare hanno costruito degli impianti di primaria importanza in zone remote, dove sarebbe possibile attaccarli solo con un numero relativamente basso di vittime.) Gli attacchi alle città potrebbero essere preceduti da un preavviso o da un ultimatum, che incoraggerebbero l'evacuazione.

In effetti, ove si eccettuino le circostanze di certe specifiche sceneggiature, ci sarà sempre l'opportunità per un paese di evacuare in una certa misura i propri centri urbani, indipendentemente dal grado di preparazione raggiunto per tale evacuazione. (Per esempio, l'1, il 2 e il 3 settembre 1939 il governo inglese evacuò circa 1 milione e mezzo di donne e di bambini dalle più importanti città dell'Inghilterra e in quegli stessi tre giorni altri due milioni di persone si trasferirono di loro iniziativa.)

Ciò nonostante, alcune infondate dichiarazioni di pretesi progetti di evacuazione in atto nell'Unione Sovietica hanno destato molta preoccupazione negli Stati Uniti. Secondo un recente rapporto della Central Intelligence Agency (CIA), se i sovietici potessero disporre di almeno una settimana per evacuare completamente le loro città e per mettere i profughi al riparo dal fallout radioattivo, le vittime della guerra potrebbero ridursi a «poche decine di milioni e i morti sarebbero una metà circa di questa cifra». Alcuni analisti sono arrivati al punto di definire «accettabili» queste perdite, tenuto conto del fatto che l'URSS lamentò 20 milioni di morti nella seconda guerra mondiale. Comunque sia, anche se l'evacuazione potesse ridurre il numero delle vittime immediate, i danni che gli USA potrebbero arrecare alle risorse economiche non proteggibili dell'URSS sarebbero così grandi che l'Unione Sovietica verrebbe cancellata come grande potenza industriale.

Uno degli scopi di qualsiasi campagna di bombardamenti strategici è quello di ridurre il potenziale di sostentamento delle forze armate nemiche schierate. Nella seconda guerra mondiale furono attaccate fabbriche, reti di trasporto e centrali elettriche. Uno degli obiettivi di queste campagne di bombardamenti era quello di distruggere le industrie su cui poggiavano altri settori economici, privando così tali settori di *input* essenziali e provocando una crescente paralisi industriale. I bombardamenti sulla Germania non raggiunsero questo obiettivo, anche a causa dei limiti delle dimensioni delle bombe a base di esplosivi chimici. Gli attacchi a una data serie di obiettivi dovevano essere distribuiti in molte incursioni, per cui gli impianti indenni potevano accogliere attrezzature di fortuna per compensare i danni subiti da certe parti della rete industriale. I civili rimasti senza casa in seguito agli attacchi aerei potevano essere ospitati in città vicine che non avevano subito danni. Anche dopo il bombardamento atomico di Hiroshima e di Nagasaki si ebbe una sufficiente disponibilità di assistenza da parte delle comunità limitrofe, in modo da poter soccorrere in maniera rilevante i sopravvissuti.

Lo spiegamento di un gran numero di armi nucleari ha modificato radicalmente il contesto dei bombardamenti strategici. Le forze attualmente schierate dagli USA e dall'URSS sono in grado di distruggere l'intera struttura industriale di qualsiasi nazione. Per di più questi danni possono essere inflitti tutti in una volta, per cui non ci sarebbe molta disponibilità di assistenza per gli obiettivi sottoposti all'attacco.

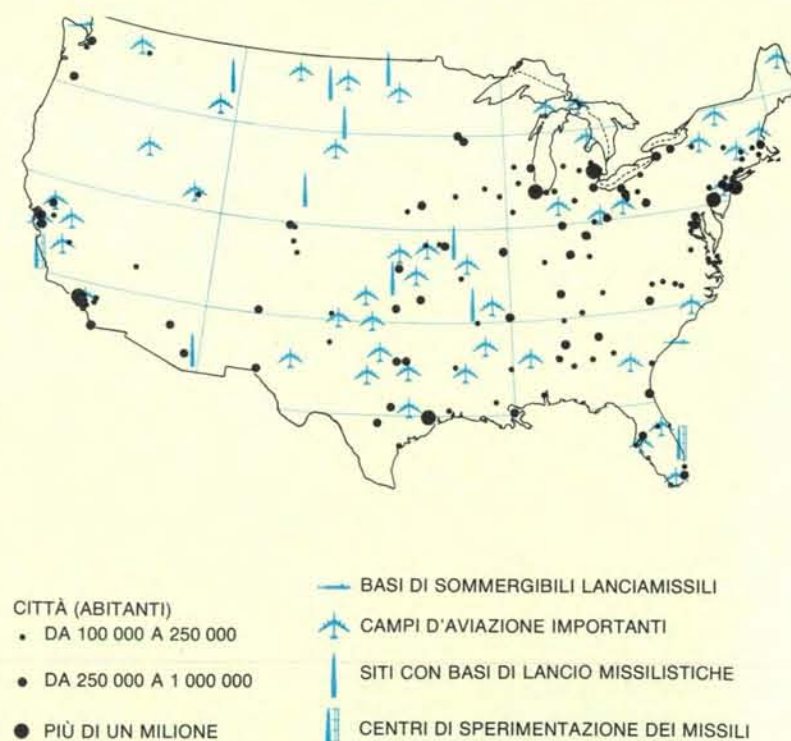
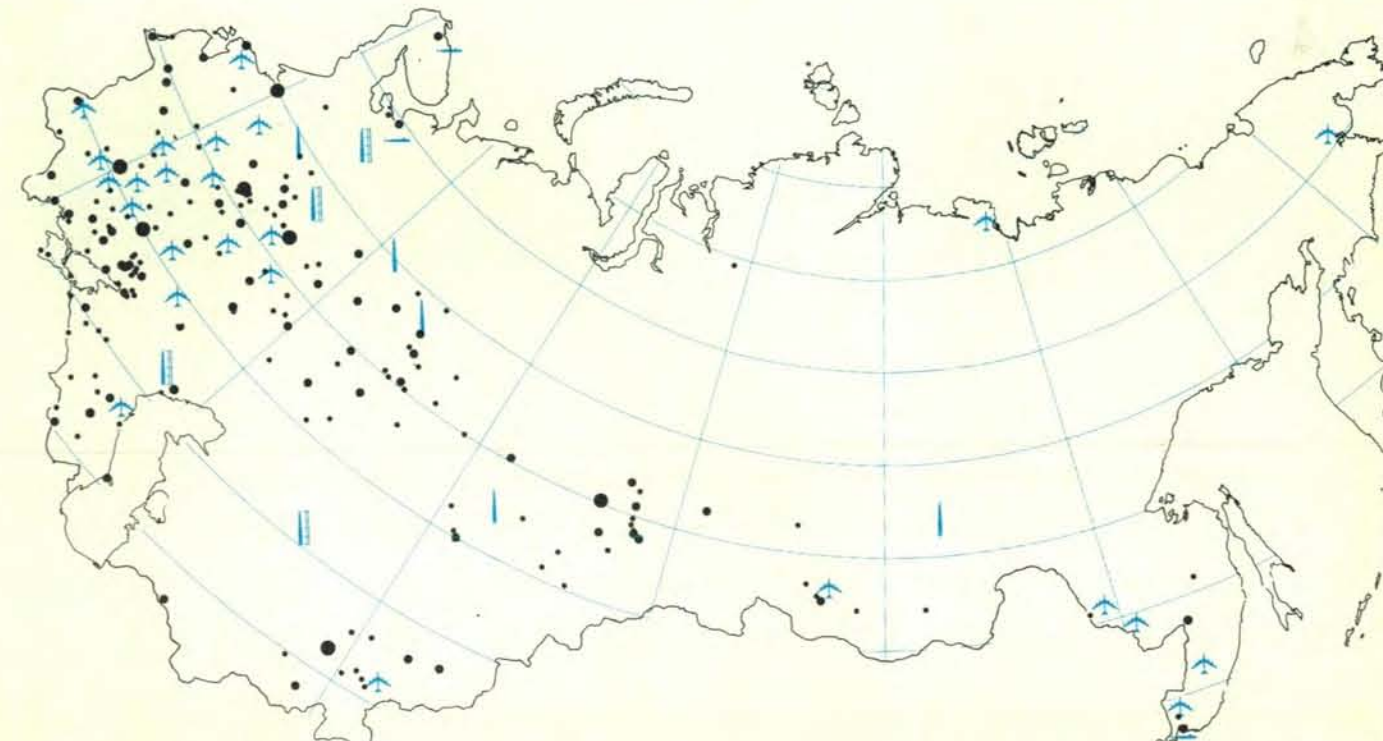
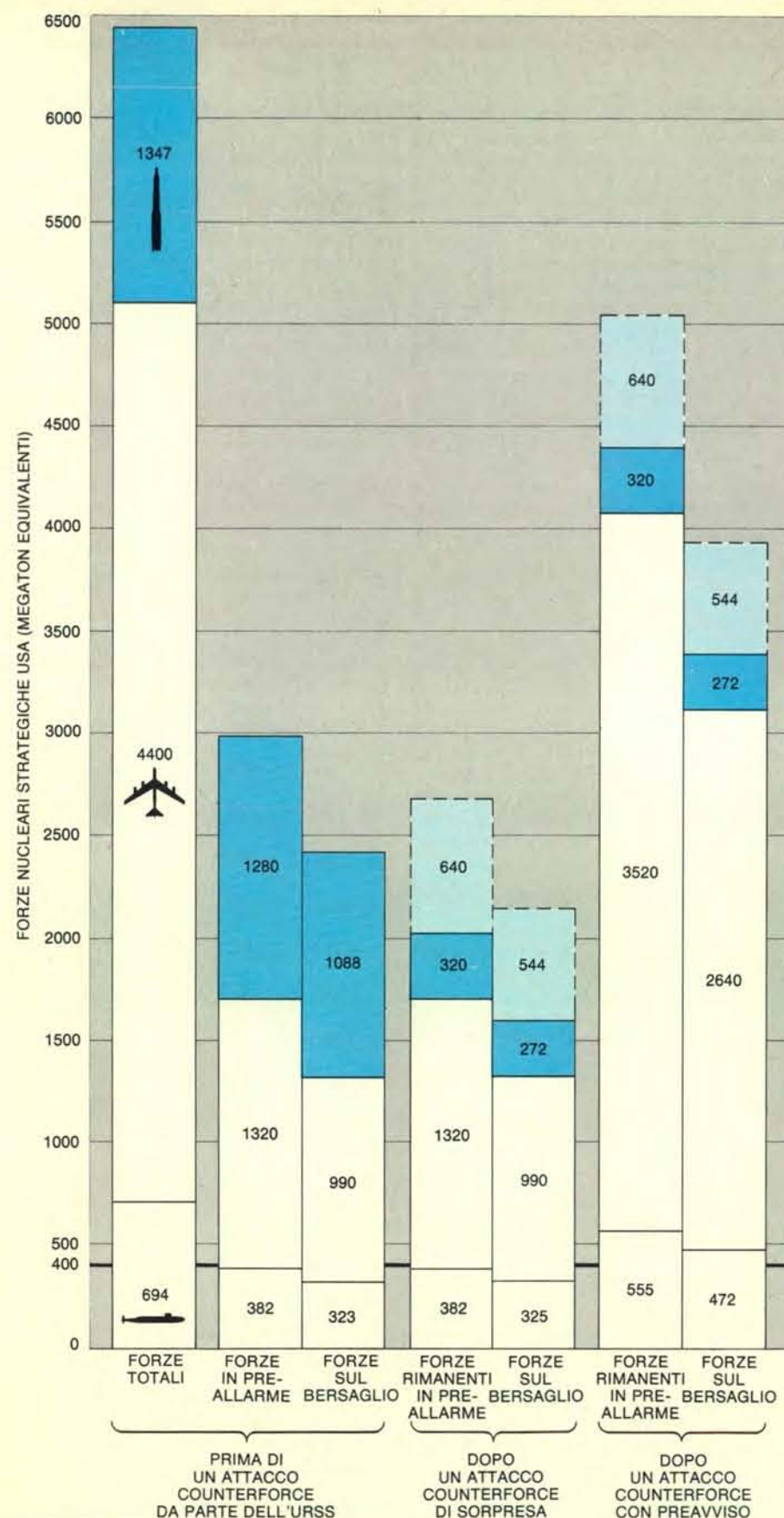
Tanto negli Stati Uniti quanto nell'Unione Sovietica un numero limitato di impianti comprende il grosso della capacità produttiva di molte industrie di primo piano. L'economia a pianificazione centralizzata dell'URSS in particolare ha molte strozzature e punti deboli vulnerabili. Basterebbe perciò la distruzione di un solo obiettivo o di pochissimi obiettivi per sconvolgere la produzione in molte altre industrie. A causa di questa concentrazione, 100 megaton equivalenti, corrispondenti alla carica esplosiva dei missili trasportati da cinque o sei sommergibili Poseidon, sarebbero sufficienti a distruggere industrie di vitale importanza senza le quali l'economia sovietica non potrebbe reggersi.

Per esempio, un recente studio dell'Office of Technology Assessment del Congresso degli Stati Uniti ha dimostrato che un attacco americano contro raffinerie di petrolio nell'URSS potrebbe, con solo 40 testate di bassa forza esplosiva, distruggere circa tre quarti di tutta la capacità di raffinazione dell'URSS. Un numero relativamente ridotto di testate potrebbe distruggere anche i trasporti, l'energia e le risorse alimentari e organizzative necessarie per qualsiasi ripresa economica postbellica. Il sistema energetico russo è particolarmente vulnerabile a un attacco, ed è di importanza vitale per la ripresa. Nell'URSS, per esempio, il trasporto interurbano di merci si svolge quasi tutto su ferrovia, mentre negli USA viene effettuato per buona parte su strada.

Bisogna stare bene attenti però a non prendere queste cifre come una prova che l'economia americana sia in qualche modo meno vulnerabile di quella sovietica. I sovietici hanno un numero di testate più che sufficiente a «coprire» analoghi obiettivi americani. In ogni caso, i piani di evacuazione, anche se avessero successo, servirebbero soltanto a rinviare, non a impedire, l'impatto della guerra sui civili. In fin dei conti, la base medica, tecnica ed educativa di una nazione uscirebbe distrutta da una guerra nucleare. Senza contare che potrebbero essere presi di mira anche le riserve di generi di prima necessità e gli altri servizi necessari per la ripresa. Anche se si riuscisse a salvare un po' di viveri, di prodotti farmaceutici, di capi di vestiario, di macchinari e di parti di ricambio, non ci sarebbero né la struttura amministrativa per distribuire le merci, né i trasporti per portarle dove ce ne fosse bisogno. La distruzione delle raffinerie e delle centrali elettriche potrebbe impedire gli approvvigionamenti e presto verrebbero a mancare certi beni. Le merci deperibili, fra cui molti prodotti alimentari e molti medicinali, andrebbero perdute se si interrompesse la fornitura di energia elettrica. La distruzione delle case renderebbe la vita difficile in estate e intollerabile in inverno. In breve, la difesa civile potrà anche proteggere qualche persona, ma non potrà certo impedire la diffusa distruzione della proprietà, essenziale per sostenere la vita. L'interdipendenza economica di una nazione industrializzata ha una vulnerabilità che non si può difendere. Anche la struttura sociale e amministrativa di una nazione sarebbe sconvolta da un attacco nucleare, al punto che un sistema politico potrebbe andare all'aria senza possibilità di ricostituzione.

Gli effetti ritardati di una guerra nucleare fra USA e URSS si propagherebbero molto al di là dei confini dei due antagonisti e dei loro alleati. Gli effetti su scala mondiale deriverebbero soprattutto dal fatto che il «fungo» della maggior parte delle esplosioni nucleari penetrerebbe nella stratosfera e vi depositerebbe materiali radioattivi di vario genere. A differenza della parte inferiore dell'atmosfera, la stratosfera manca dell'umidità e dei movimenti di scorrimento necessari per separare le sostanze gassose e quelle formate da particelle. Poiché tali materiali rimarrebbero nella stratosfera per parecchio tempo, i loro effetti verrebbero attenuati. Ma una delle conseguenze di questa lunga permanenza sarebbe un'ampia dispersione. Così, anche se meno intensi di quelli dell'atmosfera inferiore, gli effetti stratosferici durerebbero di più e sarebbero più diffusi.

Un rapporto pubblicato dalla National Academy of Science (NAS) nel 1975 elencava tre effetti della guerra nucleare che potrebbero avere un influsso sfavorevole su scala mondiale. Primo, potrebbe esaurirsi l'ozono stratosferico, poiché gli ossidi di azoto che il calore dell'esplosione nucleare produrrebbe dall'azoto e dall'ossigeno atmosferici verrebbero introdotti nella stratosfera, dove contribuirebbero alla conversione dell'ozono in ossigeno molecolare. Secondo, il deposito di grandi quantità di polvere nell'atmosfera superiore potrebbe alterare la quantità di radiazione solare che raggiunge la superficie della Terra. Terzo, pericolosi isotopi radioattivi potrebbero venire dispersi nella stratosfera e ricadere lentamente al suolo in tutto il mondo.

L'ozono stratosferico ha una funzione molto importante per la vita sulla Terra, facendo da schermo contro le dannose radiazioni ultraviolette. Il rapporto della NAS stimava che una guerra nucleare da 10 000 megaton potrebbe distruggere metà dell'ozono dell'emisfero settentrionale e circa il 30 per cento di quello dell'emisfero meridionale. Come hanno sostenuto gli oppositori degli aerei da trasporto supersonici e dei nebulizzatori al fluorocarburo, l'esaurimento dello strato di ozono potrebbe far sorgere svariati problemi di natura medica e ambientale. Potrebbero aumentare i casi di cancro e potrebbero risentirne anche le piante, comprese quelle da coltura. La distruzione dell'ozono stratosferico su questa scala

In queste due cartine si può vedere la concentrazione negli USA e nell'URSS degli obiettivi militari e civili. I puntini neri indicano l'ubicazione delle città più grandi. I simboli in colore designano le installazioni delle armi strategiche: campi d'aviazione importanti, basi di sommergibili lanciamissili, basi missilistiche a terra, siti per la sperimentazione dei missili (*si veda la legenda in basso a sinistra*). Oltre alle installazioni presentate in questa figura, gli USA dispongono di una varietà di forze strategiche di stanza in altri posti del mondo (soprattutto nell'isola di Guam e in Alaska). In generale gli obiettivi adatti a un attacco nucleare sono estremamente più concentrati nell'URSS che negli USA.

Un attacco di sorpresa, sferrato dall'URSS contro obiettivi militari strategici negli USA, lascerebbe l'arsenale americano con un numero di testate nucleari più che sufficiente a compiere una missione di distruzione sicura. Con un minimo di preavviso, il numero delle armi nucleari americane da dirigere contro obiettivi sovietici aumenterebbe in misura notevole. Il tratto spesso in nero indica i 400 megaton equivalenti sufficienti per uccidere il 35 per cento della popolazione sovietica. Le forze strategiche al di sopra di questo livello rappresentano la «sovracapacità distruttiva».

potrebbe sconvolgere la struttura termica dell'atmosfera superiore e portare a mutamenti di temperatura in tutto il mondo. È probabile che dopo una guerra di questo genere i livelli di ozono non ritornino più per molti anni a valori normali.

Una sola esplosione da un megaton a livello del suolo proietterebbe inoltre migliaia di tonnellate di polvere fine nella stratosfera. La polvere potrebbe assorbire, riflettere e disperdere le radiazioni che giungono dal Sole o che vengono riflesse dalla Terra, e si è avanzata l'ipotesi che questo effetto potrebbe provocare un cambiamento del tempo sulla superficie terrestre. Secondo lo studio della NAS, tuttavia, una guerra da 10 000 megaton non immetterebbe nella stratosfera più polvere di quanta ne fu scagliata dall'esplosione del vulcano Krakatoa nel 1883. Estrapolando da tali eventi vulcanici, il rapporto della NAS sosteneva che la cosa potrebbe risolversi soltanto in un lieve cambiamento delle condizioni del tempo in superficie.

Gli isotopi radioattivi verrebbero distribuiti in tutto il mondo dai processi di trasporto stratosferici. Dal momento che tali isotopi avrebbero un tempo di permanenza nella stratosfera relativamente lungo, molti di quelli pericolosi a vita breve decadrebbero prima di giungere al suolo. Ma alcuni isotopi pericolosi, come lo stronzio 90, il cesio 137, lo iodio 131 e il carbonio 14, continuerebbero a esistere e potrebbero entrare nelle catene alimentari della biosfera. Il rapporto della NAS non lasciava intendere che questa ricaduta radioattiva avrebbe avuto sulla vita umana tutte quelle conseguenze letali su scala mondiale che sono narrate in certi romanzi come *On the Beach* di Nevil Shute. Concentrazioni regionali di fallout nelle nazioni belligeranti (e in quelle limitrofe) potrebbero tuttavia costituire un grosso rischio radioattivo per molti evacuati e per molti residenti rurali che forse non erano stati messi in pericolo dall'attacco contro le città. «Punti caldi» meno intensi potrebbero comparire a distanze maggiori, con conseguenze biologiche sfavorevoli. Poche parti del paese assalito sfuggirebbero alla minaccia del fallout, poiché un attacco generale coprirebbe obiettivi militari ed economici in tutta la nazione, lasciando contaminate quasi tutte le zone.

I fenomeni atmosferici sono complessi e non si sa nulla sui vari modi in cui una guerra nucleare da 10 000 megaton potrebbe influire sul clima. Lo studio della NAS calcolava che gli effetti dell'esaurimento dell'ozono e dell'aggiunta di polvere non avrebbero probabilmente un influsso irreversibile sullo schema generale del tempo, ma faceva notare che non erano da escludere cambiamenti di natura molto più grave. Fra questi vari effetti non si può ignorare la possibilità di azioni sinergiche. È stato fatto notare, per esempio, che un raffreddamento generale di un solo grado centigrado potrebbe eliminare tutto il frumento che cresce nel Canada. Eventualità ancora più terribili comprendono l'espansione o la fusione del ghiaccio polare.

Il rapporto della NAS non prendeva in esame possibili cambiamenti del tempo continentale dovuti a certi effetti quali gli incendi. Una sola esplosione in aria da 10 megaton potrebbe incendiare una foresta di migliaia di chilometri quadrati. Negli Stati Uniti e nell'Unione Sovietica l'incendio di grandi praterie e foreste potrebbe privare di foglie la copertura naturale del terreno, modificando in tal modo il potere riflettente della superficie terrestre e dando origine a mutamenti del tempo. Le particelle dei prodotti della combustione gettati nell'atmosfera dagli incendi delle foreste assorbirebbero e rifletterebbero la radiazione solare e fungerebbero anche da centri di nucleazione per la formazione di goccioline d'acqua e di cristalli di ghiaccio, aumentando in tal modo la nuvolaglia e alterando la distribuzione delle precipitazioni. Questi effetti locali potrebbero aggravare i fenomeni su scala mondiale citati precedentemente.

Infine, come i vari componenti di un'economia nazionale sono collegati fra loro, così anche le nazioni sono interdipendenti. La distruzione delle economie delle maggiori potenze a opera di una guerra nucleare sarebbe un grosso colpo inferto alle economie delle nazioni che da tali potenze dipendono per lo scambio di merci e di tecnologia. I primi a soffrirne sarebbero i paesi meno sviluppati, perché in questa loro fase di sviluppo hanno un assoluto bisogno di importare tecnologia dai paesi più sviluppati.

In sintesi, gli effetti cumulativi di una guerra nucleare totale sarebbero così catastrofici da rendere priva di senso qualsiasi idea di «vittoria». Le metodologie formali delle sceneggiature di distruzione sicura non rivelano tutta l'entità di questi effetti, alcuni dei quali fra i più profondi e duraturi sono esclusi anche dalle tesi che gettano dubbi sull'adeguatezza del potere di dissuasione degli USA. Quando si considerano gli effetti ritardati di una guerra totale, dovrebbe essere chiaro a tutti che nessuna contromisura ridurrebbe in misura notevole il grado di devastazione che si avrebbe sicuramente. Anche se un programma molto efficiente di evacuazione delle città potrebbe ridurre in misura notevole le vittime immediate, esso non potrebbe impedire le conseguenze sociali ritardate della devastazione industriale ed economica. Il disastro immediato o quello ritardato sarebbero di un ordine di grandezza così enorme da escludere che l'uno o l'altro possano mai essere considerati tollerabili.

Molte iniziative sono possibili da parte degli uni o degli altri per ridurre la probabilità di una guerra nucleare totale. Molte di esse sono oggetto attualmente di un forte dissenso. Un passo nella direzione giusta sarebbe quello di riformulare in termini più realistici il concetto attualmente fuorviante di distruzione sicura, in modo da riflettere tutta l'estensione della catastrofe che sarebbe rappresentata da una guerra nucleare.

Il virus di Epstein-Barr

Sembra ormai dimostrato che esista un legame tra questo, che è il più comune dei virus che infettano l'organismo umano e che è l'agente della mononucleosi infettiva, e almeno due tipi di tumore umano

di Werner Henle, Gertrude Henle ed Evelyne T. Lennette

Da più di 70 anni è noto che certi virus sono in grado di indurre tumori negli animali in condizioni naturali o sperimentali. Già nel 1908 V. Ellerman e O. Bang in Danimarca dimostrarono che nel pollame la leucemia può essere trasmessa da un virus e nel 1911 Peyton Rous del Rockefeller Institute for Medical Research dimostrò che un tipo di sarcoma può essere trasmesso nello stesso modo nei polli. Da allora molti esempi di virus oncogeni (cioè che possono provocare tumori) sono stati scoperti nel topo, nella rana, nel coniglio, nella cavia, nella mucca, nel gatto, nel cane e persino in alcuni primati non umani. Queste scoperte hanno dato il via alla ricerca sistematica di agenti simili nell'uomo, poiché è ragionevole pensare che gli esseri umani non sfuggano al tipo di eventi che affliggono le altre specie animali.

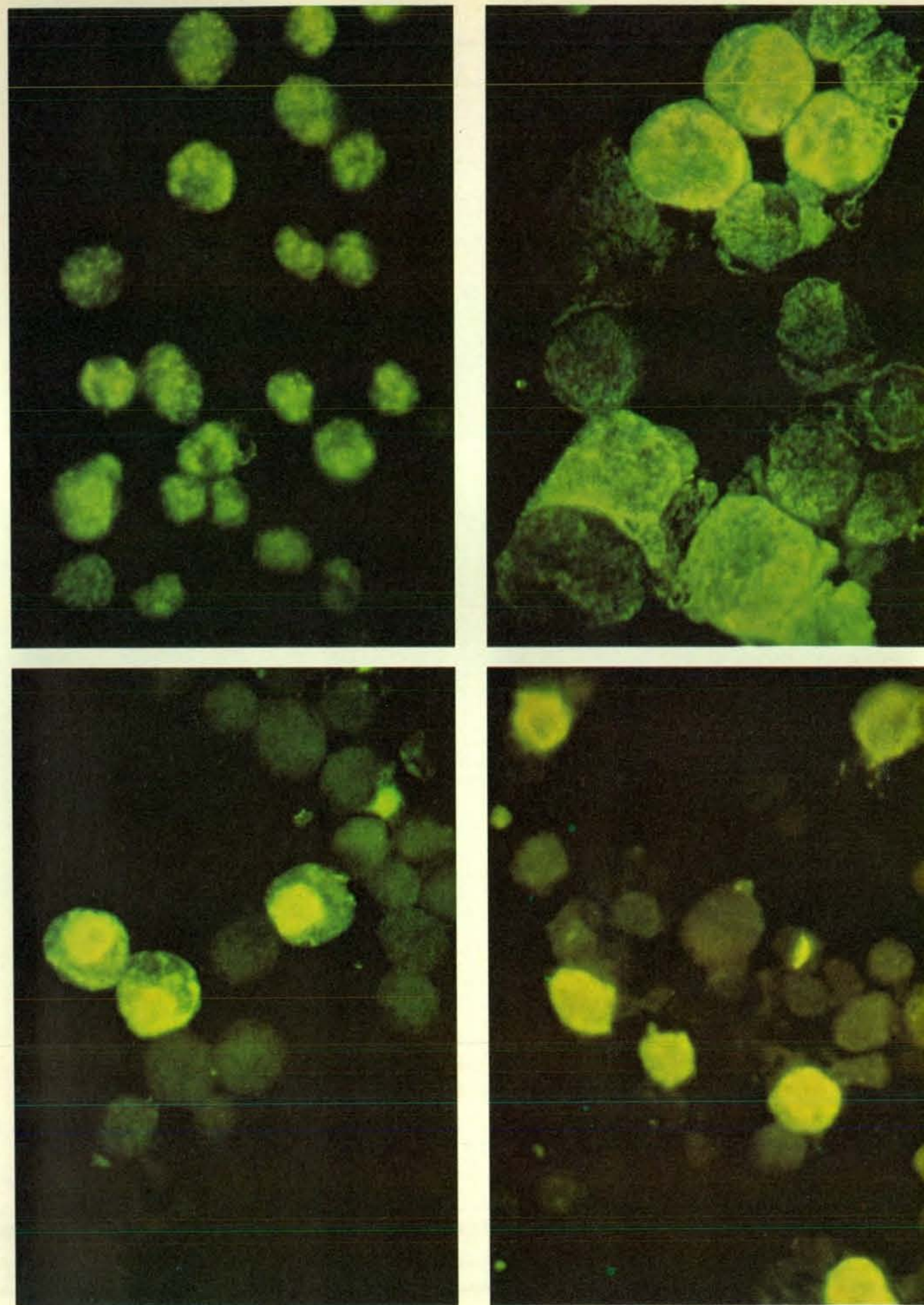
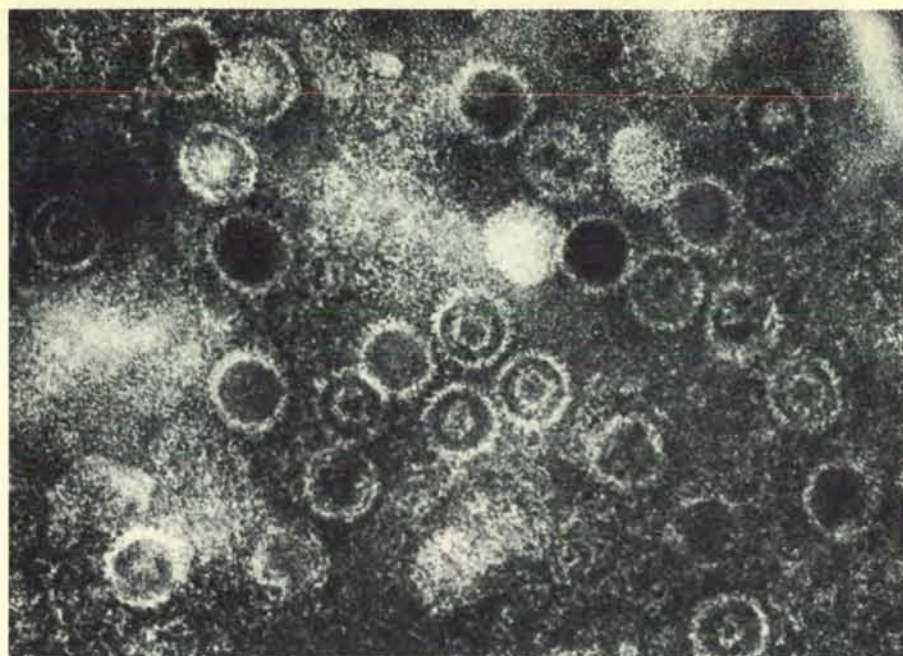
Finora il candidato più probabile come virus oncogeno dell'uomo è il virus di Epstein-Barr. L'interessante storia di questo virus cominciò verso la fine degli anni cinquanta, quando Denis Burkitt, un chirurgo inglese che lavorava all'Università di Makerere in Uganda, osservò il primo caso di un non comune tumore che avrebbe preso il suo nome. Il paziente era un ragazzo con un massiccio tumore alla mascella che si dimostrò essere un linfoma: una crescita maligna di tessuto linfoide. La crescita del tumore era straordinariamente rapida: la sua massa raddoppiava ogni 24-48 ore. Esso sfigurava in modo grottesco il paziente e provocò la sua morte a causa di metastasi in altre parti del corpo.

Burkitt ebbe presto occasione di osservare ancora la malattia, poiché gli furono portati un numero sorprendentemente grande di giovani pazienti. Il linfoma della mascella colpiva più frequentemente i bambini da sei a otto anni; era raro oltre i 14 anni e non si riscontrava in quelli di meno di un anno. Non era infrequente trovare gruppi di pazienti di entrambi i sessi a poca distanza l'uno dall'altro nello stesso gruppo di famiglie. La distribuzione del tumore ricordava quella delle malattie infettive dell'infanzia, come la poliomielite prima dell'introduzione del vaccino. Burkitt, da solo, raccolse i dati epidemiologici sul linfoma, non solo in Uganda ma in tutta l'Africa. Durante le sue spedizioni in Land Rover, che coprivano distanze di migliaia di chilometri, raccolse dati sui pazienti, la loro residenza, le loro famiglie e le precedenti malattie. Con questa massa di dati fu in grado di disegnare un profilo completo di questa forma di tumore mai precedentemente descritto.

Il linfoma di Burkitt era e rimane uno dei tumori infantili più comuni in Africa, con una frequenza di un caso ogni 10 000 bambini per anno. È riscontrabile solo in zone situate a un'altitudine inferiore ai 1800 metri sul livello del mare, una piovosità annua superiore ai 600 millimetri e una temperatura media di oltre 16 °C. Queste aree coincidono con quelle in cui le infezioni trasmesse dalle zanzare, come la malaria e la febbre gialla, sono endemiche. Burkitt provò anche che il linfoma poteva diffondersi da un'area a quelle adiacenti e che era influenzato da variazioni stagionali. Queste osservazioni lo indussero a suggerire nel 1958 che la causa del tumore fosse un agente trasmesso dalle zanzare, o che almeno un tale agente fosse un fattore scatenante della malattia.

In seguito, grazie ancora a Burkitt e ai medici che lavoravano con lui, divenne possibile curare questo linfoma letale. Già dalle prime prove si era notato che il tumore rispondeva in modo estremamente positivo alle iniezioni di ciclofosfamide, scomparendo nel giro di pochi giorni dall'inizio del trattamento. Da allora sono stati sviluppati farmaci ancora più efficaci e con minori effetti collaterali. In molti casi, tuttavia, il tumore ricompare a distanza di tempo e può dimostrarsi resistente alla chemioterapia, provocando la morte. A seconda dell'estensione iniziale e della localizzazione del tumore, tra il 30 e il 60 per cento dei pazienti sopravvive per cinque anni o più senza mostrare i sintomi della malattia.

Il linfoma di Burkitt divenne oggetto di ricerca in molti laboratori. Nel 1964 M. Anthony Epstein, B. G. Achong e Y. M. Barr del Middlesex Hospital di Londra esaminarono al microscopio elettronico le cellule ottenute dalla coltivazione in vitro di campioni di biopsie del linfoma di Burkitt. All'interno di queste cellule essi osservarono le particelle del virus che in seguito avrebbe preso il nome di virus di Epstein-Barr. Sulla base della sua struttura, il virus venne classificato come appartenente al gruppo dei virus erpetici, che hanno una doppia elica di DNA come deposito della loro informazione geneti-

Gli antigeni virali sono proteine estranee che inducono risposte anticorpali specifiche negli animali e nell'uomo. Gli antigeni del virus di Epstein-Barr possono essere identificati all'interno dei linfociti umani infettati per mezzo di sensibili tecniche che utilizzano anticorpi fluorescenti, come è mostrato nelle microfotografie della pagina a fronte. La scoperta di proteine associate al virus in due tumori umani ha fornito forti prove indiziarie per un collegamento tra il virus di Epstein-Barr e il cancro nell'uomo. La maggioranza dei linfociti infettati dal virus non produce particelle virali «figlie» ma esprime un antigene virale nel nucleo della cellula chiamato antigene nucleare di Epstein-Barr (EBNA, dalle sue iniziali inglesi). Quando gli anticorpi fluorescenti si combinano con questo antigene, il nucleo cellulare diventa brillante, ma il citoplasma rimane scuro *(in alto a sinistra)*. Prima dell'inizio del ciclo della replicazione virale nella cellula produttrice infettata appaiono gli antigeni precoci. Il primo antigene, il componente diffuso *(D)*, si trova sia nel nucleo che nel citoplasma *(in alto a destra)*. Il secondo, il componente ristretto *(R)*, è confinato in una grossa massa nel citoplasma *(in basso a sinistra)*. Dopo la replicazione del DNA virale compaiono gli antigeni tardivi; questi antigeni sono costituenti proteici del virus stesso. Uno degli antigeni tardivi è l'antigene del capside virale (VCA), che è presente sia nel nucleo che nel citoplasma *(in basso a destra)*. Ognuna delle tre malattie sicuramente associate al virus di Epstein-Barr dà origine a una caratteristica produzione di anticorpi specifici contro gli antigeni virali.

Le particelle di virus di Epstein-Barr sono ingrandite circa 96 000 volte in questa fotografia al microscopio elettronico eseguita da M. Anthony Epstein della University of Bristol Medical School. I capsidi, o gusci proteici, delle particelle virali sono distinguibili contro lo sfondo scuro. In alcuni di essi è penetrato del materiale che fa da contrasto: è così possibile osservare che essi sono vuoti o contengono il DNA virale. Le particelle virali non contrastate appaiono bianche. Il virus è stato isolato da due tumori umani: il linfoma di Burkitt e il carcinoma nasofaringeo.

ca. A quel tempo erano noti tre virus umani di questo tipo: il virus dell'*herpes simplex*, che causa vescicole sulla pelle e le mucose (le cosiddette febbri); il citomegalovirus, che induce serie malformazioni congenite e un'ampia varietà di sintomi nei pazienti più anziani, e l'*herpes zoster*, che provoca la varicella e il fuoco di Sant'Antonio. La scoperta di un virus erpetico nelle cellule di linfoma di Burkitt non destò eccessivo interesse, tuttavia, poiché a quel tempo i virus erpetici non erano indiziati come agenti oncogeni degli animali. Così si suppose che le particelle virali nelle cellule tumorali fossero dei semplici ospiti occasionali.

Mentre Epstein tentava invano di interessare i virologi inglesi all'identificazione del virus, noi stavamo cercando di ottenere alcuni campioni di linfoma di Burkitt da pazienti africani per il nostro laboratorio del Children's Hospital di Filadelfia. Una serie di eventi fortuiti e fortunati portarono Epstein in contatto con noi e, nel giro di pochi giorni, parte della sua collezione di colture cellulari attraversava l'Atlantico e approdava al nostro laboratorio. Poco dopo l'arrivo delle colture fummo in grado di dimostrare che il virus di Epstein-Barr era un nuovo tipo di virus erpetico. Giungemmo a questa conclusione in seguito all'impossibilità di trasmettere il virus a varie colture cellulari standard, agli embrioni di pulcino e agli animali di laboratorio suscettibili all'infezione da parte dei tre virus erpetici noti.

Convinti che si trattasse di un virus sconosciuto in precedenza, incominciammo a studiare i rapporti tra il virus di Epstein-Barr e il linfoma di Burkitt. Cercammo dapprima gli anticorpi contro il virus nel sangue dei pazienti. Quando qualcuno è stato esposto a un virus, quasi invariabilmente produce anticorpi specifici contro di esso che rimangono in circolo. La presenza nel sangue di anticorpi contro un dato virus è quindi un segno inequivocabile che nel passato il paziente è stato esposto a quel virus.

Per determinare se il virus di Epstein-Barr giocasse un ruolo nella genesi del linfoma di Burkitt cercammo di rispondere alle seguenti domande: tutti i pazienti affetti dal linfoma di Burkitt possiedono gli anticorpi contro il virus di Epstein-Barr? Gli anticorpi contro il virus sono limitati ai pazienti affetti dal linfoma o si trovano anche in individui sani o in pazienti con altre malattie? Nel caso che gli anticorpi non siano limitati al linfoma di Burkitt, i pazienti affetti dal linfoma mostrano un titolo (concentrazione nel plasma sanguigno) significativamente più alto o una maggiore varietà di anticorpi contro il virus di Epstein-Barr rispetto agli altri individui?

Diagramma della struttura del virus di Epstein-Barr appartenente al gruppo dei virus erpetici, i quali immagazzinano la loro informazione genetica sotto forma di una doppia elica di DNA. Le particelle virali infettive consistono di tre componenti: una massa centrale a ciambella, il nucleoide, che contiene il DNA virale in forma condensata; il capside icosaedrico, costituito di subunità proteiche tubulari cave, chiamate capsomeri; la guaina protettiva che deriva o dalla membrana nucleare o dal plasmalemma della cellula ospite. La guaina incorpora un certo numero di proteine virali (antigeni) che sono state prodotte e inserite nella membrana prima dell'inizio del montaggio finale delle particelle virali. Durante l'infezione la guaina virale si fonde con il plasmalemma di una cellula ospite suscettibile all'infezione e il nucleocapside (il capside e il DNA virale contenuto in esso) entra nella cellula. In seguito, il DNA virale può persistere allo stato represso o può influire sui meccanismi enzimatici della cellula ospite inducendo la produzione di nuove particelle virali.

Per individuare gli anticorpi contro il virus di Epstein-Barr utilizzammo la sensibile tecnica nota come immunofluorescenza indiretta. Le cellule in coltura che contengono il virus vengono fissate su un vetrino, su cui successivamente viene depositato un po' di siero del paziente che si ritiene possa essere dotato di anticorpi contro il virus. In questo modo gli anticorpi del paziente possono legarsi agli antigeni (proteine estranee) del virus sia sulla superficie sia all'interno della cellula. Le cellule vengono poi lavate e incubate con una seconda serie di anticorpi che reagiscono specificamente con tutti gli anticorpi umani e sono stati accoppiati con le molecole di un colorante fluorescente. Quando le cellule vengono illuminate con la radiazione ultravioletta, al microscopio gli anticorpi del paziente si riconoscono come aree fluorescenti all'interno delle cellule infettate dal virus. La fluorescenza è intensa perché molti anticorpi fluorescenti possono legarsi a un singolo anticorpo del paziente.

Con questo metodo semplice, ma elegante fummo in grado di confermare che il virus di Epstein-Barr è in effetti un nuovo tipo di virus, dato che nessuno degli anticorpi contro i virus erpetici noti reagisce con le cellule del linfoma di Burkitt. Poi ci procurammo il siero di un ragazzo nigeriano affetto da linfoma di Burkitt che era stato mandato al Clinical Center of the National Institutes of Health a Bethesda, nel Maryland. Nel test dell'immunofluorescenza indiretta, il siero del ragazzo provocò l'intensa fluorescenza di un piccolo numero di cellule di linfoma di Burkitt. Quando le singole cellule fluorescenti vennero osservate al microscopio, esse mostrarono la presenza di numerose particelle del virus di Epstein-Barr; le cellule non fluorescenti non contenevano le particelle virali.

Con nostra grande sorpresa riscontrammo gli anticorpi contro il virus di Epstein-Barr non solo nei bambini affetti da linfoma di Burkitt, ma anche nel siero della quasi totalità dei bambini africani sani. Ancora più inaspettata fu la scoperta che gli anticorpi contro il virus sono presenti nei bambini di molte parti del mondo, il che implica che il virus di Epstein-Barr ha una diffusione su scala mondiale e che praticamente nessuno sfugge all'infezione. Divenne presto evidente che l'età in cui si acquisiscono gli anticorpi dipende dalle condizioni di vita prevalenti. Nelle condizioni di affollamento e scarsa igiene che prevalgono nei paesi in via di sviluppo praticamente tutti i bambini producono anticorpi contro il virus a partire dall'età di tre o quattro anni. Nei paesi più avanzati l'acquisizione degli anticorpi contro il virus è spesso rimandata fino all'adolescenza o alla maturità.

Il titolo dell'anticorpo contro il virus di Epstein-Barr nei bambini affetti da linfoma di Burkitt era da otto a dieci volte superiore a quello dei bambini sani. I pazienti affetti da linfoma reagivano quindi al virus in modo anormale. Tuttavia trovammo che anche pazienti con altre malattie hanno alti titoli di anticorpi contro il virus di Epstein-Barr. In particolare i pazienti affetti da carcinoma nasofaringeo avevano titoli anticorpali 10 volte più alti degli individui sani o di altri pazienti affetti da tumori diversi della testa e del collo. Il collegamento tra questo carcinoma e il virus di Epstein-Barr venne stabilito per la prima volta da Lloyd G. Old e i suoi colleghi al Memorial Sloan-Kettering Cancer Center, i quali osservarono che il siero dei pazienti affetti da vari tipi di carcinoma nasofaringeo forma dei precipitati con gli estratti di cellule di linfoma di Burkitt. Questa scoperta implicava che il virus di Epstein-Barr è intimamente associato a entrambe le malattie.

La distribuzione del linfoma di Burkitt in Africa rivela che si trovano gruppi di casi *(cerchi neri)* nelle aree in cui la malaria è endemica a causa di fattori climatici *(in colore)*. Il linfoma di Burkitt è un tumore della mascella e di altri tessuti che colpisce i bambini africani ed è strettamente associato al virus di Epstein-Barr. Questo dato suggerisce che la malaria possa essere un cofattore nella genesi del linfoma. La cartina è stata disegnata in base al lavoro di Denis Burkitt e altri.

Il carcinoma nasofaringeo è un tumore maligno che si origina dal tessuto epiteliale che tappezza le cavità retronasali. Colpisce in prevalenza gli adulti ed è particolarmente diffuso (10 persone su 100 000 ogni anno) nella Cina meridionale, ma si presenta anche, con frequenza più ridotta, in Tunisia e nell'Africa orientale. Anche gli eschimesi che vivono in Alaska vanno soggetti a questo male, che invece è raro nel resto del mondo. Il tumore di solito segue uno di questi decorsi letali: può crescere attraverso le ossa del cranio e invadere il cervello, oppure le cellule del carcinoma possono diffondersi dal tumore primario attraverso i vasi linfatici, formando metastasi nei linfonodi del collo e successivamente nel fegato, nei polmoni, in altri organi e nelle ossa.

Il fatto che gli anticorpi contro il virus di Epstein-Barr fossero ubiquitari nelle popolazioni sane ci fece sospettare che il virus fosse la causa di qualche malattia comune e non particolarmente pericolosa, oltre a essere in qualche modo associato ai due tumori maligni. Per una curiosa coincidenza l'identità di questa malattia fu svelata da una scoperta fortuita fatta nel nostro laboratorio. Verso la fine del 1967 una dei nostri giovani tecnici di laboratorio accusò i sintomi classici della mononucleosi infettiva: mal di gola, febbre e gonfiore dei linfonodi del collo. Prima della malattia nel suo siero non erano presenti gli anticorpi contro il virus di Epstein-Barr, ma questi comparvero in coincidenza con l'inizio della malattia. Questo evento fortuito fornì il primo indizio che il virus di Epstein-Barr potesse essere la causa lungamente cercata della mononucleosi infettiva, che era stata descritta per la prima volta da T. P. Sprunt e F. A. Evans della John Hopkins University School of Medicine nel 1920.

Nella sua forma classica la mononucleosi infettiva causa febbre, tonsillite, gonfiore dei linfonodi e ingrossamento del fegato e della milza. Malgrado il suo nome non sembra essere particolarmente infettiva; non sono mai stati riportati casi di grosse epidemie. La maggior parte dei pazienti è costituita da giovani adulti e i casi che non presentano complicazioni si risolvono con la completa guarigione in un periodo di tempo che va da una a quattro settimane. Meno dell'uno per cento dei pazienti presenta complicazioni (come encefaliti, polmoniti e collassi renali); i casi mortali sono molto rari. Alcuni clinici, d'altra parte, hanno descritto la mononucleosi come un fenomeno autolimitato di leucemia. I pazienti colpiti da questa malattia mostrano un netto aumento del numero dei globuli bianchi, che spesso sono anche morfologicamente anormali. I globuli bianchi in eccesso includono i linfociti B (che si originano dal

midollo osseo e producono anticorpi) e i linfociti T (che maturano nel timo e partecipano alle reazioni immunitarie cellulari). Nei pazienti affetti da mononucleosi la proliferazione anomala dei linfociti ha presto termine; nella leucemia invece le cellule continuano a moltiplicarsi senza controllo.

Per poter confermare la relazione causale tra il virus di Epstein-Barr e la mononucleosi avevamo bisogno di altri campioni di siero di pazienti colpiti da questa malattia. Ci ricordammo che James C. Niederman e Robert W. McCollum della Yale University School of Medicine erano impegnati dal 1958 in una indagine a lungo termine sulla mononucleosi negli studenti universitari: essi avevano raccolto il siero di centinaia di studenti prima (al momento dell'iscrizione all'università), durante e dopo la malattia. Essi ci misero a disposizione la loro preziosa collezione perché potessimo eseguire le nostre prove. Trovammo che gli anticorpi contro il virus di Epstein-Barr, assenti prima della malattia, comparivano durante e dopo di essa. Questo ci permise di concludere inequivocabilmente che il virus di Epstein-Barr è responsabile di almeno una malattia: la mononucleosi infettiva. Studi epidemiologici successivi hanno indicato che chi è stato colpito dalla mononucleosi infettiva non ha una maggiore probabilità di contrarre il cancro in un periodo successivo della sua vita. Ciò è in accordo col fatto che praticamente ognuno di noi viene infettato dal virus di Epstein-Barr, anche coloro i quali non presentano i sintomi della mononucleosi.

Sono stati fatti molti progressi nella comprensione di come il virus di Epstein-Barr causi la mononucleosi. È ora noto che la via di infezione negli individui non immuni comincia dalla bocca e dal tratto respiratorio, a causa dell'uso di bicchieri e posate contaminate, oppure di contatti orali (di qui il nome familiare di «malattia del bacio») o dell'inalazione di particelle virali. Il bersaglio iniziale del virus è costituito probabilmente dalle cellule epiteliali della bocca e della gola, ma il virus prima o poi infetta i linfociti B, che lo disseminano nel sistema linfatico. (Per qualche motivo sconosciuto il virus di Epstein-Barr è altamente selettivo per i linfociti B e non infetta i linfociti T.) I primi sintomi della malattia si manifestano di solito da quattro a sette settimane dopo l'infezione; durante la fase acuta della malattia si possono riscontrare particelle virali nella saliva di quasi tutti i pazienti.

Nel corso della sua evoluzione il virus di Epstein-Barr ha stabilito una simbiosi quasi perfetta con il suo ospite umano, permettendo la sopravvivenza di entrambi. Benché l'infezione virale stabilisca un'immunità permanente, il virus persiste in forma latente nei linfociti B del sistema linfatico e in alcune cellule viene attivato spontaneamente. La progenie di particelle virali infettive viene perciò riversata periodicamente nella saliva, assicurando in questo modo la diffusione del virus agli individui ancora non immuni. Nelle condizioni igieniche primitive prevalenti nei tempi preistorici e in molte parti del mondo ancora oggi l'infezione primaria da parte del virus di Epstein-Barr ha luogo nei primi anni di vita, quando non causa malesseri significativi, forse a causa dell'immaturità delle cellule bersaglio o del sistema immunitario. Le migliori condizioni igieniche e la migliore qualità della vita dei paesi sviluppati ritardano spesso le infezioni primarie del virus verso la fine dell'infanzia e l'adolescenza, quando possono più facilmente provocare i sintomi della mononucleosi. Per esempio, nella popolazione universitaria degli Stati Uniti, tra il 35 e l'80 per cento degli studenti possiedono gli anticorpi contro il virus di Epstein-Barr al momento dell'iscrizione, mentre tra il 10 e il 15 per cento di quelli che non possiedono gli anticorpi vengono infettati ogni anno. Circa i due terzi di questi individui infettati tardivamente accusano disturbi di una qualche entità. Questo spiega perché la mononucleosi è diffusa essenzialmente solo nei paesi economicamente avanzati.

Il virus di Epstein-Barr assomiglia ad altri virus erpetici per il fatto che può vivere allo stato latente, o represso; esso persiste nelle cellule linfoidi umane senza indurre la produzione di particelle virali «figlie» e senza provocare la morte della cellula ospite. I linfociti B umani coltivati in vitro possono venire classificati in due gruppi: le linee cellulari «produttrici» sono quelle in cui il genoma virale (ossia l'insieme completo dei geni virali) può venire attivato spontaneamente in una piccola percentuale delle cellule (fino al 10 per cento), dando origine agli antigeni virali e occasionalmente a particelle virali infettanti. Le linee cellulari «non produttrici» sono quelle in cui il genoma virale non si attiva mai spontaneamente. Anche quando il genoma virale di una cellula non produttrice viene attivato sperimentalmente, il processo è di solito abortivo: possono venire prodotti gli antigeni virali ma non le particelle virali complete.

La presenza dei genomi virali nelle cellule non produttrici fu dimostrata per la prima volta da Harald zur Hausen e Heinrich Schulte-Holthausen dell'Università di Erlangen. Essi dimostrarono che le eliche di DNA virale estratte dalle particelle virali infettive si appaiano alle eliche complementari di DNA estratto dalle cellule non produttrici. Fu dimostrato che le cellule infettate in maniera latente possono contenere fino a 50 copie complete del genoma virale.

Alice Adams e Tomas Lindahl collaboratori di George Klein al Karolinska Institutet di Stoccolma scoprirono che solo una parte del DNA virale può venire separato dal DNA della cellula ospite mediante centrifugazione. La frazione separabile consiste di plasmidi circolari di DNA virale, della stessa lunghezza delle eliche di DNA impacchettate nelle particelle virali infettanti. La frazione non separabile è costituita da copie del DNA virale che si sono inserite linearmente nel DNA della cellula ospite, dove vengono replicate insieme al resto dei cromosomi dell'ospite.

Di tanto in tanto in una cellula della linea cellulare produttrice il DNA virale cessa di essere represso: di conseguenza inizia l'attiva produzione di particelle virali e la cellula muore. Le particelle virali riversate nel terreno di coltura da queste cellule produttrici sono di due tipi: trasformanti e litiche. Il virus litico induce la produzione di particelle virali figlie nelle cellule infettate. Il virus trasformante, prodotto dalla maggioranza delle cellule produttrici, provoca la moltiplicazione indefinita dei linfociti B quando questi vengono coltivati in vitro. (I linfociti B che non sono stati infettati non possono crescere e riprodursi in coltura e generalmente muoiono nel giro di due o tre settimane). Se il virus trasformante viene inattivato dal calore o dal trattamento con anticorpi o viene rimosso dal terreno di coltura facendolo passare attraverso un filtro a pori estremamente fini, i linfociti B non vengono trasformati. Questi esperimenti hanno stabilito inequivocabilmente che l'agente trasformante è il virus stesso.

Il virus di Epstein-Barr trasforma i linfociti anche negli essere umani viventi: i linfociti B ottenuti dal sangue e dai linfonodi degli individui infettati dal virus crescono facilmente in coltura, ma quelli degli individui non infettati non crescono. Giovanni Rocchi e i suoi colleghi dell'Università di Roma hanno dimostrato che negli individui sani con infezioni latenti del virus di Epstein-Barr un linfocita su 10 milioni è trasformato; nei pazienti affetti da mononucleosi infettiva è trasformato un linfocita su 5000. Questi risultati suggeriscono che il virus di Epstein-Barr sia potenzialmente cancerogeno poiché conferisce ai linfociti B la capacità di crescere e riprodursi indefinitamente, una delle proprietà caratteristiche delle cellule tumorali. La trasformazione dei linfociti sembra essere indotta dalle copie del DNA virale che sono state inserite linearmente nel cromosoma dell'ospite. Elliott Kieff della University of Chicago School of Medicine sta ora analizzando in dettaglio il DNA virale nel tentativo di identificare il particolare segmento di DNA che porta l'informazione per l'attività trasformante del virus.

Benché la maggior parte del DNA virale nelle cellule non produttrici sia repressa, come è indicato dall'assenza della replicazione virale, una piccola frazione del DNA virale viene espressa nella produzione di nuove proteine, sia nel nucleo cellulare, sia nella membrana cellulare. Queste proteine non sono proteine strutturali del virus trasformante, ma sono codificate da esso. Dato che le proteine associate al virus differiscono dalle normali proteine della cellula, inducono la produzione di anticorpi. Uno di questi antigeni è l'antigene nucleare di Epstein-Barr (EBNA, dalle sue iniziali inglesi), che è localizzato nel nucleo della cellula in associazione con la cromatina, il materiale che incorpora il DNA. Esso è stato scoperto da Beverly Reedman e Klein, i quali hanno anche dimostrato che l'EBNA è espresso da tutte le cellule che portano il genoma virale, comprese le cellule del linfoma di Burkitt e del carcinoma nasofaringeo. Nelle cellule non produttrici la presenza dell'EBNA nel nucleo è la sola indicazione che la cellula è stata trasformata dal virus di Epstein-Barr.

Modificando le condizioni di crescita in coltura dei linfociti infettati in modo latente o trattandoli con determinati farmaci, è possibile scatenare nelle cellule il ciclo litico, che culmina nella morte o nella lisi (distruzione) della cellula. Il ciclo litico implica la produzione di molti altri antigeni e, se procede a sufficienza, la liberazione delle particelle virali figlie. Appena dopo l'induzione del ciclo litico e prima che sia possibile riscontrare la replicazione del DNA virale, nella cellula appaiono tre antigeni specifici del virus. Questi antigeni vengono chiamati complesso antigenico precoce. Il primo antigene, il componente diffuso *(D)*, compare inizialmente nel nucleo cellulare e si diffonde in tutto il citoplasma. Il secondo antigene, il componente ristretto *(R)*, è localizzato in una grossa massa nel citoplasma adiacente al nucleo cellulare. Il terzo antigene, l'antigene precoce di membrana, è inserito nella membrana della cellula ospite. In molte cellule il processo infettivo si ferma a questo stadio.

In altre cellule, tuttavia, il ciclo litico procede con la replicazione del DNA vi-

L'induzione spontanea della replicazione del virus di Epstein-Barr può aver luogo nei discendenti di una singola cellula di una coltura che produce cellule di linfoma di Burkitt. Molte cellule contenenti il genoma virale (l'insieme completo dei geni del virus) persistono indefinitamente, dividendosi ogni 18-24 ore. In una percentuale delle cellule che arriva al 10 per cento, tuttavia, il genoma virale represso viene attivato, causando la produzione di nuove particelle virali. La loro liberazione è inevitabilmente accompagnata dalla morte e dalla lisi delle cellule ospiti.

La molecola di DNA che porta l'informazione genetica del virus di Epstein-Barr è presente in molte copie nei linfociti umani infettati sia in modo latente che in modo produttivo. La maggior parte delle molecole di DNA sono in forma circolare, come è mostrato qui, e si replicano indipendentemente dai cromosomi della cellula ospite. Il resto delle molecole di DNA virale è integrato linearmente nel DNA della cellula ospite e viene copiato insieme a esso. Il DNA all'interno della particella virale è in forma lineare. Questa microfotografia è stata eseguita da G. Bjursel dell'Università di Aarhus.

rale. Successivamente a questo stadio appaiono gli antigeni virali tardivi: queste proteine sono i componenti strutturali del capside (il guscio proteico) della particella virale. Gli antigeni tardivi comprendono l'antigene del capside virale (VCA) e gli antigeni di membrana tardivi, che sono inseriti nella membrana cellulare. Le particelle virali figlie vengono montate nel nucleo cellulare, dopo di che alcune di esse passano nel citoplasma della cellula e vengono estruse attraverso la membrana cellulare. Durante questo processo i capsidi virali vengono ricoperti da una guaina protettiva derivata dalle membrane della cellula ospite. La liberazione delle particelle virali figlie, o anche la sintesi pura e semplice degli antigeni virali precoci, porta inevitabilmente alla morte della cellula ospite.

La sequenza degli eventi che seguono l'infezione da parte del virus di Epstein-Barr nell'uomo è assai meno chiara, perché il periodo di incubazione tra l'esposizione al virus e la comparsa dei sintomi clinici della mononucleosi dura da 30 a 40 giorni. Per seguire il decorso dell'infezione è utile tenere sotto controllo la risposta immunitaria del paziente. Nell'infezione virale primaria compare per prima l'immunoglobulina M (IgM), che è specifica per il VCA. Durante il periodo di incubazione di quattro-sette settimane la produzione di IgM si interrompe e viene sostituita dalla produzione di immunoglobulina G (IgG). Nel momento in cui il paziente si rivolge al medico gli anticorpi IgM e IgG contro il VCA hanno di solito raggiunto i loro livelli massimi.

Più dell'80 per cento dei pazienti affetti da mononucleosi produce anche anticorpi temporanei IgG contro il componente *D* del complesso antigenico precoce nel momento in cui la malattia si manifesta. Al contrario, gli anticorpi contro l'EBNA si sviluppano tardi nel corso della malattia e poi persistono per tutta la durata della vita.

I pazienti affetti da mononucleosi infettiva sviluppano anche anticorpi eterofili, così chiamati perché reagiscono con antigeni non correlati di diverse specie animali. Per esempio, gli anticorpi eterofili umani agglutinano (formano dei precipitati con) i globuli rossi di pecora o cavallo. Gli anticorpi eterofili furono scoperti nel 1932 da John R. Paul e W. W. Bunnel della Yale School of Medicine e si dimostrarono altamente specifici per la mononucleosi. La determinazione di questi anticorpi nel plasma dei pazienti è ancora il metodo diagnostico principale per la mononucleosi a causa della semplicità dell'analisi. Per quanto ne sappiamo, però, l'antigene responsabile dell'anticorpo eterofilo non è collegato direttamente al virus di Epstein-Barr e non tutti i pazienti producono questi anticorpi.

Come abbiamo già detto, i bambini dei paesi in via di sviluppo vengono esposti al virus di Epstein-Barr molto precocemente. Recenti studi su neonati del Ghana, eseguiti da noi in collaborazione con Robert J. Biggar dei National Institutes of Health, hanno dimostrato che gli anticorpi materni contro il virus passano dalla madre al feto attraverso la placenta e proteggono dall'infezione il neonato fino a sei mesi dopo la nascita. Nei 12 mesi successivi, comunque, l'80 per cento dei bambini da noi studiati produsse i propri anticorpi contro il virus senza manifestare alcun segno clinico della malattia. La risposta anticorpale dei bambini piccoli al virus di Epstein-Barr è simile alla reazione degli adulti alla mononucleosi, con due notevoli eccezioni: gli anticorpi eterofili sono o assenti o presenti a livelli insignificanti, e gli anticorpi IgG temporanei contro il complesso antigenico precoce sono diretti contro il componente *R* anziché contro il componente *D*. (Gli anticorpi contro il componente *D* sono identificabili in modo caratteristico nei pazienti affetti da linfoma di Burkitt.) Il significato di questa differenza nei profili anticorpali nei bambini piccoli e negli adulti non è ancora stato compreso.

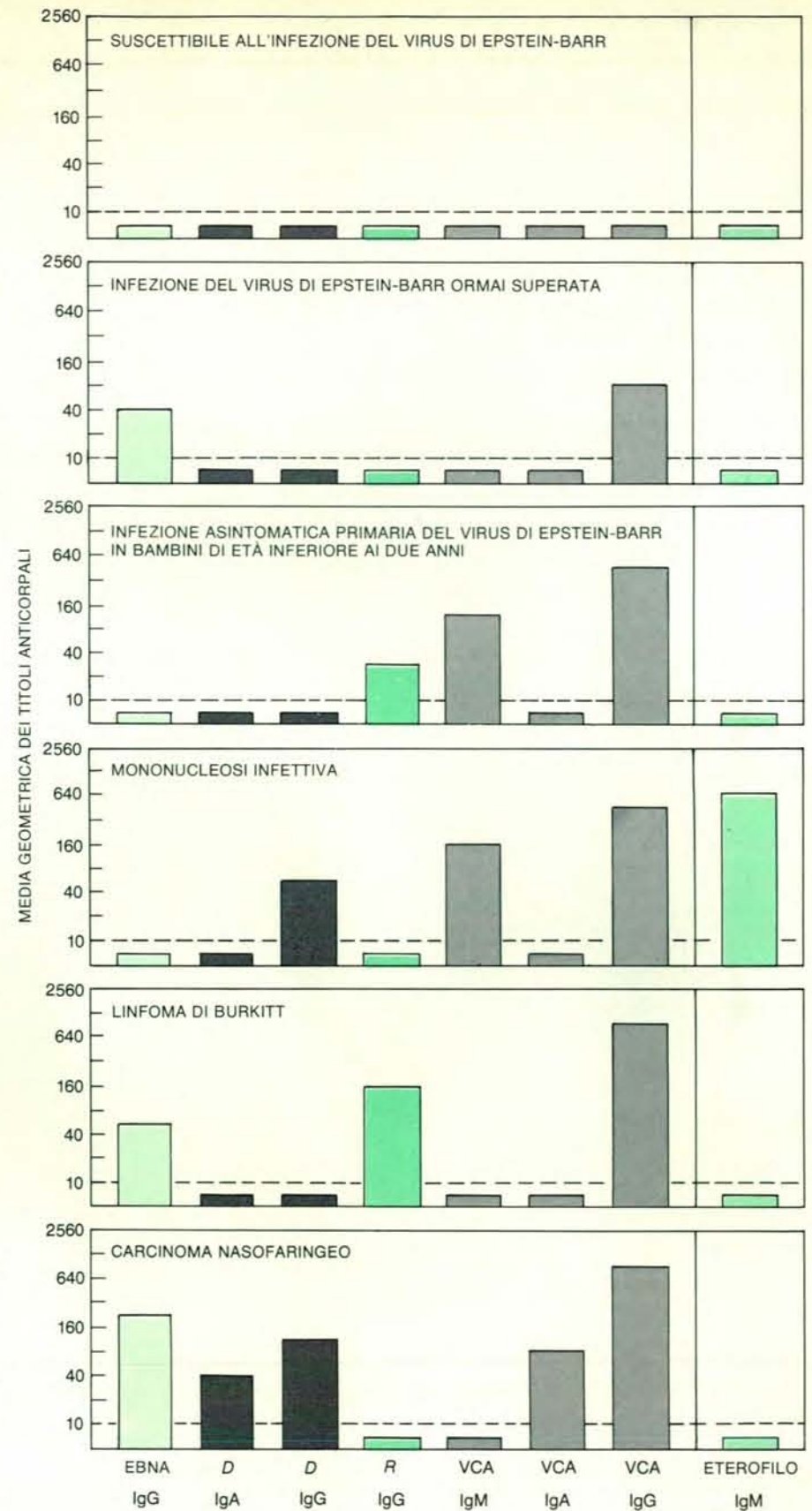
VIRUS DI EPSTEIN-BARR
CELLULE NON PRODUTTRICI
CELLULE PRODUTTRICI
TRASFORMAZIONE
EBNA
ANTIGENI VIRALI E PARTICELLE VIRALI «FIGLIE»
ULTERIORI DIVISIONI
VIRUS TRASFORMANTE
RISPOSTA ANTICORPALE PRECOCE
VIRUS LITICO
PARZIALE DISTRUZIONE
CELLULE SOPRAVVISSUTE
RILASCIO DI EBNA
CELLULE SOPRAVVISSUTE
DISTRUZIONE PARZIALE
INDUZIONE SPONTANEA
RISPOSTA ANTICORPALE TARDIVA
ULTERIORI DIVISIONI
RILASCIO DI EBNA
ANTIGENI VIRALI
EQUILIBRIO NELLA PRODUZIONE DI ANTICORPI

Questa diagramma riassume la sequenza degli eventi che si verificano durante la mononucleosi infettiva dell'uomo. Molti linfociti B vengono infettati liticamente dal virus di Epstein-Barr durante la diffusione dell'infezione; essi producono antigeni virali e particelle virali prima della loro inevitabile morte e lisi. Gli antigeni virali inducono una risposta anticorpale precoce. Le particelle virali figlie sono principalmente del tipo trasformante; i linfociti B infettati rimangono vivi e si dividono indefinitamente. Le cellule trasformate contengono copie del DNA virale ed esprimono sia l'antigene nucleare di Epstein-Barr sia un antigene sulla superficie della cellula che diventa il bersaglio dei linfociti T attivati. La successiva distruzione dei linfociti B trasformati da parte dei linfociti T provoca alcuni dei più noti sintomi clinici della mononucleosi; libera anche abbastanza EBNA da indurre una risposta anticorpale tardiva contro questo antigene. Le cellule trasformate non vengono mai eliminate completamente. Alcune delle cellule sopravvissute vengono indotte spontaneamente alla produzione di particelle virali. Quelle che, invece, vengono distrutte liberano gli antigeni e portano l'organismo a un livello di equilibrio nella produzione di anticorpi.

Mentre le prove che collegano il virus di Epstein-Barr alla mononucleosi sono inconfutabili, il ruolo giocato dal virus nei due tipi di cancro non è così evidente. Finora gli argomenti a favore di una relazione causale tra il virus di Epstein-Barr e il linfoma di Burkitt sono i seguenti: 1) è possibile riscontrare il DNA o gli antigeni virali nel tumore, 2) il virus trasforma i linfociti B in coltura, 3) il virus induce linfomi in primati non umani, 4) i titoli anticorpali contro il virus sono più alti nei pazienti affetti da linfoma rispetto a quelli di una popolazione controllo, 5) è possibile correlare i profili anticorpali e la prognosi del tumore. Questi cinque criteri hanno tutti fornito un valido sostegno all'ipotesi di una relazione causale tra il virus di Epstein-Barr e il linfoma di Burkitt, sebbene rimangano ancora aperte molte importanti questioni.

Il DNA del virus di Epstein-Barr è stato trovato nella grande maggioranza (98 per cento) delle biopsie del linfoma di Burkitt in Africa, ma non in altri tipi di linfomi. Ogni cellula tumorale contiene molte copie del genoma virale. Le cellule infettate liticamente che producono gli antigeni precoci, il VCA o le particelle virali figlie non si osservano nelle biopsie fresche, ma compaiono uno o due giorni dopo che le cellule sono state messe in coltura. È probabile che le cellule infettate liticamente compaiano in piccolo numero all'interno del tumore, ma vengano rapidamente distrutte dalle difese immunitarie del paziente non appena esprimono l'antigene di membrana precoce sulla superficie della cellula. In effetti già dal 1966 Klein e i suoi colleghi hanno riscontrato un antigene sulle cellule del linfoma di Burkitt e hanno scoperto che spesso esso è ricoperto dagli anticorpi del paziente stesso. Il virus di Epstein-Barr è quindi intimamente associato alle cellule del linfoma di Burkitt.

Le colture di linfociti B trasformate dal virus di Epstein-Barr tuttavia differiscono per alcuni importanti aspetti dalle colture di cellule di linfoma di Burkitt, che derivano anch'esse dai linfociti B. Mentre le colture di linfociti trasformati dal virus derivano da un certo numero di cellule diverse, le colture di linfoma di Burkitt sono monoclonali: ogni tumore deriva da una singola cellula trasformata in modo maligno. Inoltre la grande maggioranza delle cellule di linfoma di Burkitt mostrano una anormalità cromosomica che non si riscontra nei linfociti trasformati o nei linfociti dei pazienti affetti da mononucleosi. Questa aberrazione cromosomica, scoperta da G. Manolov, Y. Manolova e i loro colleghi del laboratorio di A. Levan presso l'Università di Lund, consiste nella traslocazione di un piccolo segmento di uno dei cromosomi numero 8 sul braccio lungo di uno dei cromosomi numero 14, e di un piccolo segmento del cromosoma 14 sul cromosoma 8. Aberrazioni del cromosoma 14 simili, ma non identiche, sono state scoperte in altri linfomi.

L'origine e la funzione dell'aberrazione cromosomica del linfoma di Burkitt non sono note, ma sembra che ciò conferisca un vantaggio biologico alle cellule interessate. Qualunque siano le sue cause, l'aberrazione potrebbe essere un importante fattore o cofattore che predispone all'insorgenza del linfoma di Burkitt, dato che i linfociti dei pazienti affetti da mononucleosi non possiedono cromosomi

I profili anticorpali prodotti in risposta al virus di Epstein-Barr differiscono significativamente nelle varie malattie associate al virus. Sono stati identificati profili caratteristici per le infezioni primarie latenti dei bambini di età inferiore ai due anni, per la mononucleosi dei giovani adulti, per il linfoma di Burkitt e per il carcinoma nasofaringeo. I pazienti possono produrre anticorpi di tre diverse classi di immunoglobuline (IgM, IgA e IgG), che vengono diretti contro i seguenti antigeni associati al virus: l'antigene nucleare di Epstein-Barr, l'antigene del capside virale e i componenti diffuso e ristretto del complesso antigenico precoce. Gli anticorpi contro l'antigene eterofilo sono presenti in quantità notevole esclusivamente nei pazienti affetti da mononucleosi infettiva.

aberranti. Ancora una volta, tuttavia, la situazione non è del tutto chiara: le traslocazioni sono osservabili in molti, ma non in tutti i linfomi di Burkitt associati al virus di Epstein-Barr, e si trovano anche in alcune cellule di linfoma di Burkitt non associate al virus. Sembra probabile che la trasformazione dei linfociti da parte del virus di Epstein-Barr costituisca solamente il primo passo nel processo di trasformazione maligna e che siano necessari altri eventi per rendere le cellule completamente maligne.

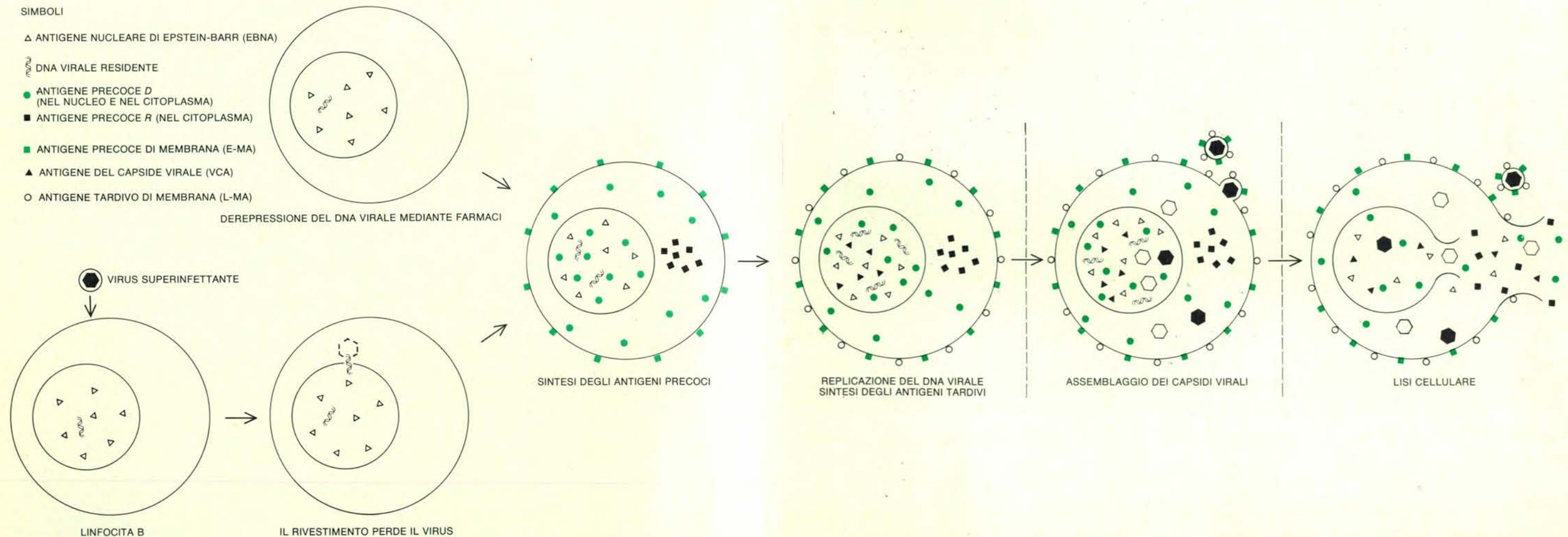
Gli esperimenti sui primati subumani hanno rafforzato l'associazione del virus di Epstein-Barr con il linfoma di Burkitt. Thomas Shope, D. C. De-Chairo e I. George Miller della Yale School of Medicine hanno iniettato il virus trasformante nella callitrice e in altre scimmie del Nuovo Mondo e hanno osservato un certo numero di risposte differenti. Alcuni animali producevano gli anticorpi contro il virus senza manifestare alcun sintomo di malattia, altri mostravano una proliferazione temporanea dei linfociti e i rimanenti venivano colpiti da linfomi letali tra le sei e le 14 settimane seguenti. Le cellule del linfoma esprimevano l'EBNA e gli animali producevano gli anticorpi contro il VCA e il complesso antigenico precoce. I linfomi indotti dal virus di Epstein-Barr nelle scimmie non sono però del tutto assimilabili al linfoma di Burkitt dell'uomo: sono diffusi anziché solidi e si sviluppano nel giro di alcune settimane invece che richiedere anni come spesso avviene nel caso dell'insorgenza dell'infezione primaria dei bambini africani e la successiva comparsa del linfoma di Burkitt. Questi esperimenti hanno comunque dimostrato chiaramente che il virus di Epstein-Barr può indurre lo sviluppo di linfomi maligni.

I pazienti affetti da linfoma di Burkitt hanno generalmente un alto titolo anticorpale contro il VCA e il componente *R* del complesso antigenico precoce; nello stadio finale della malattia possono venire prodotti anticorpi contro il componente *D*. Il controllo costante e prolungato nel tempo dei pazienti ha rivelato che coloro i quali inizialmente rispondono bene alla terapia, ma mantengono titoli alti o crescenti degli anticorpi contro il VCA o l'antigene *R* generalmente vanno incontro a ripetute e fatali recidive del tumore, anche dopo periodi di apparente scomparsa del tumore che durano da tre a sei anni. Al contrario, i pazienti che, dopo l'intervento chirurgico o la chemioterapia, mostrano una diminuzione costante dei titoli anticorpali contro l'antigene *R* hanno una buona probabilità di sopravvivere a lungo senza mostrare i sintomi della malattia. Gli anticorpi contro il virus di Epstein-Barr sono quindi degli ottimi indicatori della prognosi del linfoma di Burkitt.

Sebbene praticamente tutti i bambini africani vengano infettati dal virus di Epstein-Barr entro i primi tre anni di vita, l'incidenza massima del linfoma di Burkitt cade in corrispondenza dei sei-otto anni. Per spiegare questa interessante discrepanza nel 1972 fu organizzato sotto gli auspici dell'International Agency for Research on Cancer uno studio prospettico del linfoma nel West Nile District in Uganda, una regione dove questa malattia è endemica. Si raccolsero campioni di sangue di più di 45 000 bambini dagli uno agli otto anni, e il plasma fu conservato per eventuali analisi successive. Tra il 1972 e il 1977, 14 di questi bambini furono colpiti dal linfoma di Burkitt. Il loro plasma raccolto precedentemente fu allora analizzato per riscontrare gli anticorpi contro il virus di Epstein-Barr e fu confrontato con il plasma di bambini sani di età e sesso corrispondenti appartenenti a famiglie che abitavano nei dintorni e furono effettuati numerosi altri controlli.

Si scoprì che i bambini ammalati avevano acquisito gli anticorpi contro il virus da 7 a 54 mesi prima che il tumore fosse diagnosticato e che i loro titoli anticorpali contro il VCA erano superiori ai titoli di quasi tutti i controlli. Non era quindi strano che la comparsa del tumore non fosse accompagnata da un aumento significativo degli anticorpi contro il VCA rispetto al livello antecedente alla malattia. Tuttavia 10 dei 14 pazienti acquisirono nuovi anticorpi contro l'antigene *R*. Questi risultati esclusero la possibilità che il linfoma di Burkitt fosse la conseguenza immediata di una rara infezione primaria da parte del virus di Epstein-Barr in bambini africani in età relativamente avanzata. La nostra conclusione fu che un intenso e persistente stato di portatore virale, indicato dagli alti livelli degli anticorpi, predispone i bambini al linfoma di Burkitt, ma che lo sviluppo del tumore richiede ulteriori fattori. La malaria, che è altamente endemica nelle regioni in cui il linfoma di Burkitt è prevalente, potrebbe essere uno di questi fattori. Una forte infezione di parassiti malarici potrebbe predisporre i bambini al linfoma di Burkitt perché è noto che la malaria sopprime il sistema immunitario e stimola la proliferazione dei linfociti. È ora in fase di realizzazione uno studio epidemiologico dell'effetto dei programmi di lotta alla malaria sull'incidenza del linfoma di Burkitt: questo dovrebbe fare un po' di luce sull'apparente collegamento tra le due malattie. Un altro importante fattore nella genesi del linfoma potrebbe essere rappresentato dall'età in cui avviene l'infezione primaria da parte del virus di Epstein-Barr. Le infezioni nella prima fanciullezza potrebbero trasformare i linfociti B immaturi, permettendo loro di persistere nel corpo e di dare origine alla fine alle cellule del linfoma. Ulteriori studi su queste infezioni precoci potrebbero quindi fornire risultati interessanti.

Benché il linfoma di Burkitt sia endemico in molte regioni dell'Africa orientale e occidentale, esso si manifesta sporadicamente in altre parti del mondo, inclusi gli Stati Uniti. Tuttavia circa l'80 per cento dei casi americani non sono associati al virus di Epstein-Barr: nei tumori non si riscontra alcuna cellula contenente il DNA virale o EBNA-positiva, e i pazienti hanno profili anticorpali simili a quelli dei bambini normali. Il 20 per cento residuo dei casi americani di linfoma di Burkitt sono associati al virus di Epstein-Barr e sono indistinguibili dalle loro controparti africane, a parte il fatto che negli Stati Uniti l'intervallo di età dei pazienti è spostato verso l'adolescenza e la giovinezza. Di recente nel giro di un solo anno è stato osservato nelle campagne della Pennsylvania un gruppo di quattro giovani adulti affetti da linfoma di Burkitt associato al virus. I pazienti vivevano entro 30 miglia l'uno dall'altro, e due di essi, che erano sposati, avevano vissuto insieme per molti anni.

I casi di linfoma di Burkitt non associati al virus di Epstein-Barr (non solo l'80 per cento dei casi americani ma anche il 2 per cento circa di quelli africani) costituiscono chiaramente un dilemma: o il linfoma associato al virus è una malattia differente dal linfoma non associato al virus (sebbene i tumori siano morfologicamente indistinguibili) o il linfoma di Burkitt ha almeno due cause, una delle quali è il virus di Epstein-Barr. In ogni caso, il fattore secondario che si ritiene indispensabile per l'insorgenza del tumore non può esse-

Il ciclo di replicazione litica del virus di Epstein-Barr può essere indotto in una cellula di linfoma di Burkitt in coltura o mediante una seconda infezione del virus o dereprimendo il genoma virale con antimetaboliti. Quando è iniziata la sintesi del complesso antigenico precoce, la produzione delle proteine cellulari si ferma e la cellula è destinata a morire, sebbene il processo possa arrestarsi a diversi stadi *(linee tratteggiate)*. La maggior parte delle cellule esprime gli antigeni precoci, altre esprimono gli antigeni tardivi e poche liberano particelle virali.

re la malaria nei casi americani di linfoma di Burkitt associato al virus. D'altra parte non è stato individuato alcun altro fattore secondario e, nei casi di linfoma di Burkitt non associati al virus di Epstein-Barr, non è stato trovato alcun altro virus.

Gli argomenti in favore dell'associazione tra il virus di Epstein-Barr e il carcinoma nasofaringeo sono basati sugli stessi criteri che suggeriscono l'associazione tra il virus e il linfoma di Burkitt. Finora tutte le biopsie del carcinoma eseguite in tutte le parti del mondo hanno rivelato la presenza del DNA virale, che è assente in tutti gli altri carcinomi della testa e del collo.

Molti esperimenti hanno dimostrato che il DNA virale si trova nelle cellule del carcinoma e non nei linfociti che si infiltrano nel tumore. Per esempio, le impronte dei frammenti di biopsia sui vetrini contengono cellule di carcinoma depositate, nelle quali è possibile riscontrare l'EBNA. In questi tumori non sono state trovate cellule infettate liticamente, presumibilmente perché queste cellule vengono rapidamente distrutte dalle difese immunitarie del paziente. Tuttavia, dopo che le cellule di carcinoma sono state messe in coltura, il genoma virale può essere attivato per mezzo di alcuni farmaci, dando inizio alla produzione dell'antigene *D*, del VCA e anche a particelle virali infettive che sono capaci di trasformare i linfociti B in coltura. D'altra parte, non è stato possibile trasformare col virus di Epstein-Barr alcuna cellula epiteliale normale coltivata in vitro, e l'iniezione del virus in primati non umani non ha indotto alcun carcinoma. Probabilmente la ragione di questi insuccessi è che sono state utilizzate metodiche sbagliate nella somministrazione del virus oppure che sono richiesti cofattori ignoti.

I pazienti affetti dal carcinoma nasofaringeo mostrano una risposta anticorpale caratteristica contro il virus di Epstein--Barr. A differenza dei pazienti affetti dal linfoma di Burkitt essi producono un alto titolo di anticorpo contro il componente *D* del complesso antigenico precoce, ma non contro il componente *R*. Un'altra caratteristica singolare del carcinoma è la produzione di anticorpi della classe delle immunoglobuline A (IgA). Anticorpi di questo tipo contro il VCA e l'antigene *D* sono frequentemente presenti e raggiungono alti livelli.

In uno studio in collaborazione con John C. Ho del Queen Elizabeth Hospital di Hong Kong abbiamo dimostrato che nei pazienti affetti da carcinoma nasofaringeo i titoli degli anticorpi IgG e IgA contro il VCA e l'antigene *D* salgono in concomitanza dell'aumento in volume del tumore e poi declinano se la terapia è stata efficace. In alcuni pazienti ci fu un iniziale abbassamento dei titoli per un periodo di due-quattro anni, dopo di che la tendenza si invertì e il titolo ricominciò a salire. In seguito questi pazienti furono colpiti da una recidiva del tumore o da metastasi. Il controllo costante dei titoli anticorpali contro il virus di Epstein-Barr costituisce quindi un mezzo per valutare l'efficacia della terapia del carcinoma nasofaringeo e per prevedere eventuali recidive del tumore.

Queste osservazioni suggeriscono che il virus di Epstein-Barr giochi un ruolo causale nell'insorgenza del carcinoma nasofaringeo, ma come nel caso del linfoma di Burkitt rimangono ancora parecchie perplessità. Benché anche nelle regioni in cui il carcinoma è endemico l'infezione primaria da parte del virus di Epstein-Barr avvenga in una fase molto precoce della vita, il tumore colpisce in Cina solamente gli adulti, con una incidenza massima verso i 40 anni. In Tunisia la distribuzione delle età è bimodale, con circa il 20 per cento dei casi che riguardano soggetti di età inferiore ai 20 anni. Il lungo intervallo tra l'infezione virale primaria e l'insorgenza del tumore di nuovo porta a pensare alla partecipazione di cofattori. Tra i cinesi sembra evidente l'esistenza di una predisposizione genetica al tumore: in alcune famiglie si registrano molti casi nel giro di tre o quattro generazioni. I pazienti cinesi affetti dal carcinoma mostrano anche con alta frequenza un certo antigene di istocompatibilità (HLA), ma questo antigene non si osserva nei pazienti tunisini. Un altro fattore facilitante potrebbe essere l'ingestione o l'inalazione di sostanze cancerogene come le nitrosammine, che sono presenti in vari alimenti conservati tra cui il pesce affumicato e nel fumo del fuoco di legna e dell'incenso.

In conclusione, il virus di Epstein-Barr è uno dei virus umani più ampiamente distribuiti. È sicuramente la causa della mononucleosi infettiva e tutte le prove disponibili puntano verso un suo coinvolgimento nella genesi del linfoma di Burkitt e del carcinoma nasofaringeo. Entrambi questi tumori derivano dalla divisione incontrollata di una singola cellula trasformata in modo maligno e questo significa che il genoma del virus di Epstein-Barr doveva essere presente in quella prima cellula. Sembra improbabile che la vera causa dei tumori, qualunque essa sia, possa sempre scegliere senza motivi precisi le cellule che portano il genoma virale. Tali cellule, dopo tutto, rappresentano solo una frazione estremamente piccola della popolazione totale di linfociti B da cui ha origine il linfoma di Burkitt. Considerazioni simili valgono anche per i precursori delle cellule del carcinoma. È perciò improbabile che il virus di Epstein--Barr sia semplicemente un ospite innocuo di questi tumori: se non è il fattore principale, è per lo meno una delle condizioni necessarie.

Sembra chiaro che sono necessari parecchi eventi per rendere le cellule trasformate dal virus completamente maligne. Questi eventi potrebbero essere l'azione di cancerogeni di natura chimica, fisica o biologica. Una grave malnutrizione, l'azione concomitante di malattie come la malaria, un abbassamento delle difese immunitarie dell'ospite o uno disequilibrio di una qualsiasi delle varie funzioni dell'organismo possono costituire ulteriori fattori facilitanti.

Nei bambini africani le infezioni primarie da parte del virus di Epstein-Barr avvengono poco dopo la scomparsa degli anticorpi materni contro il virus, che erano stati trasmessi al feto attraverso la placenta. In questo grafico viene indicato il momento dell'infezione virale primaria in relazione al decadimento degli anticorpi materni e all'età di insorgenza di 512 casi di linfoma di Burkitt. L'incidenza massima del tumore viene raggiunta parecchi anni dopo l'infezione virale primaria. Questo fa ritenere che la genesi del tumore coinvolga altri fattori oltre al virus.

La presenza di un antigene virale associato al virus di Epstein-Barr in una biopsia di carcinoma nasofaringeo dà credito al collegamento tra il virus e la genesi di questo tumore nell'uomo. Su vetrini per microscopio sono state ottenute delle impronte di campioni di biopsie. Le impronte sono state colorate con anticorpi fluorescenti contro l'antigene nucleare di Epstein-Barr *(in alto)* e con siero mancante degli anticorpi contro l'antigene *(in basso)*. I nuclei delle cellule maligne sono brillanti in presenza dell'anticorpo, ma rimangono scuri in sua assenza. Questa tecnica è stata sviluppata dagli autori e da Dolly Huang e John C. Ho del Queen Elizabeth Hospital di Hong Kong.

Il vaccino contro il virus di Epstein--Barr, oltre a essere utile per la prevenzione della mononucleosi, potrebbe fornire una ulteriore conferma per il ruolo causale del virus nel linfoma di Burkitt e nel carcinoma nasofaringeo se prevenisse la loro insorgenza nelle aree in cui questi tumori sono endemici. Tuttavia lo sviluppo di tale vaccino è disseminato di ostacoli che potrebbero rivelarsi insormontabili. Per prima cosa, il virus non può essere ancora prodotto in grande quantità e finora può essere ricavato solo da cellule dotate di caratteri maligni. Inoltre i vaccini costituiti da virus erpetici uccisi si sono generalmente dimostrati insoddisfacenti oppure hanno fornito una protezione solo temporanea. I vaccini costituiti da virus vivi potrebbero essere utilizzati se il virus potesse essere attenuato, ma tale attenuazione potrebbe aumentare il suo potenziale cancerogeno in modo per ora imprevedibile.

C'è un altro problema: nelle regioni in cui il linfoma di Burkitt o il carcinoma nasofaringeo sono endemici il vaccino dovrebbe essere somministrato durante il breve e incerto periodo dopo che dal sangue del neonato sono scomparsi gli anticorpi materni contro il virus di Epstein--Barr (che vengono trasmessi al feto attraverso la placenta) e prima che il neonato venga infettato naturalmente. Gli anticorpi materni impedirebbero al vaccino di indurre la produzione di nuovi anticorpi da parte del sistema immunitario del neonato e l'infezione naturale da parte del virus potrebbe avvenire appena dopo la scomparsa degli anticorpi materni. Questa questione della scelta dei tempi presenta un problema logistico che sarebbe difficile risolvere nelle condizioni ambientali prevalenti nella regione africana interessata dal linfoma di Burkitt. Tuttavia è necessario tentare di superare questi ostacoli tecnici, se non altro per la prevenzione su scala mondiale della mononucleosi infettiva.

Questo resoconto serve a illustrare le difficoltà che si presentano nel fornire la prova definitiva del coinvolgimento dei virus nell'induzione dei tumori umani. Malgrado il fatto che tutti i criteri indiretti per collegare un virus a un tumore umano siano stati soddisfatti dal virus di Epstein--Barr, il ruolo esatto del virus nell'insorgenza di questi due tumori rimane incerto. In effetti, una prova finale che convinca tutti potrebbe non essere mai raggiunta. Per chi è disposto ad accettare l'evidenza indiretta, tuttavia, il virus di Epstein-Barr è il candidato più probabile per il ruolo di primo virus oncogeno umano riconosciuto.

Generazione di microonde mediante diodo Gunn

Questo dispositivo, di apparente semplicità, racchiude un gran numero di effetti legati agli elettroni «caldi» ed è in grado di generare medie potenze a microonde con tensioni di alimentazione poco elevate

di Fiorenzo Ardemagni

Nel campo della generazione di microonde, i componenti elettronici convenzionali hanno sempre trovato un limitato impiego a causa del rapido deterioramento delle loro prestazioni al crescere della frequenza. Lo stesso transistore, nonostante siano trascorsi ormai trent'anni dalla sua invenzione, ha guadagnato terreno faticosamente, tant'è che solo in questi ultimi tre o quattro anni si è affacciato alle soglie dei 10 GHz (dieci miliardi di cicli al secondo).

Conseguentemente, il settore degli oscillatori ad altissima frequenza si è evoluto in maniera pressoché indipendente sviluppando dispositivi adeguati alle necessità, basati tutti sulle proprietà di fasci di elettroni nel vuoto spinto, fra i quali, probabilmente, i più noti sono il *Magnetron* e il *Klystron*. Questi speciali tubi elettronici sono assai complessi dal punto di vista costruttivo e richiedono tensioni di alimentazione molto elevate (dell'ordine del migliaio di volt). Al contrario, il diodo Gunn è un piccolo dispositivo semiconduttore a due terminali in grado di generare medie potenze a microonde con bassa rumorosità; oltre alla semplicità costruttiva esso ha l'enorme vantaggio di funzionare con tensioni di alimentazione dell'ordine della decina di volt.

Lo sviluppo di questo componente elettronico ebbe inizio nel 1963 con la sensazionale osservazione di J. B. Gunn, destinata a dare un nuovo impulso alla fisica dei semiconduttori e, di riflesso, alla elettronica dello stato solido. Mentre compiva, ai laboratori della International Business Machines, ricerche sulle proprietà del rumore nei semiconduttori, Gunn notò che, sottoponendo una barretta di arseniuro di gallio (GaAs) o di fosfuro di indio (InP), a intensi campi elettrici statici, di entità variabile fra 2000 e 4000 volt per centimetro, la corrente nel circuito, invece di assumere un valore ben definito congruente alla tensione applicata, oscillava ad alta frequenza *(si veda l'illustrazione in alto nella pagina a fronte)*. Nel corso delle prove, vennero misurati valori di frequenza compresi fra 0,47 e 6,50 GHz, dipendenti solo dal campione impiegato.

Gunn non riuscì a spiegare il meccanismo responsabile delle oscillazioni, e si limitò a denunciare la presenza del fenomeno sperimentale. Solo l'anno successivo H. Kroemer puntualizzò, in una breve nota, che la maggior parte delle proprietà osservate dell'effetto Gunn potevano essere spiegate, almeno qualitativamente, assumendo per valida la teoria sviluppata, tra gli anni 1961 e 1962, da B. K. Ridley e T. B. Watkins e, indipendentemente, da C. Hilsum. Per comprendere pienamente gli aspetti principali di tale teoria, è necessario premettere alcuni risultati fondamentali della fisica dei solidi e, in particolare, dei semiconduttori.

Un solido cristallino ideale è caratterizzato da una disposizione regolare tridimensionale di atomi, ottenuta grazie alla ripetizione nello spazio di una cella elementare. Nel caso dell'arseniuro di gallio, gli atomi di arsenico e di gallio formano due reticoli cubici a facce centrate che si compenetrano secondo lo schema tipico della zincoblenda. Nel cristallo, la conducibilità elettrica è legata alle possibilità di movimento degli elettroni al suo interno. I nuclei, infatti, sono fissi, salvo i moti di vibrazione compiuti, per effetto termico, nell'intorno delle loro posizioni di equilibrio.

Per interpretare il comportamento che gli elettroni e, in generale, tutte le particelle elementari manifestano sotto certe condizioni, è indispensabile ipotizzare una loro natura ondulatoria «compenetrata», se così si può dire, con quella più familiare di corpuscoli materiali. Le caratteristiche dell'apparato sperimentale che viene impiegato per rivelarne la presenza fanno sì che, di volta in volta, possa essere messo in luce un aspetto di questa duplice natura.

Nel caso dell'elettrone «libero», il cui moto avvenga senza che vi siano perturbazioni da parte di campi di forze esterne, l'onda associata all'elettrone ha una frequenza angolare e un numero d'onda k (numero d'onde vibrazionali per unità di lunghezza) direttamente proporzionali all'energia e alla quantità di moto della particella, rispettivamente. La costante di proporzionalità è $\hbar = h/2\pi$, essendo h la costante di Plank.

La relazione esistente tra ω e k viene detta *curva di dispersione* ed è di fondamentale importanza nella fisica dei fenomeni ondulatori. Nel caso dell'elettrone libero, l'energia è direttamente proporzionale al quadrato del momento lineare (quantità di moto) tramite l'inverso della massa della particella. Pertanto, la curva di dispersione ω *(k)* sarà una parabola.

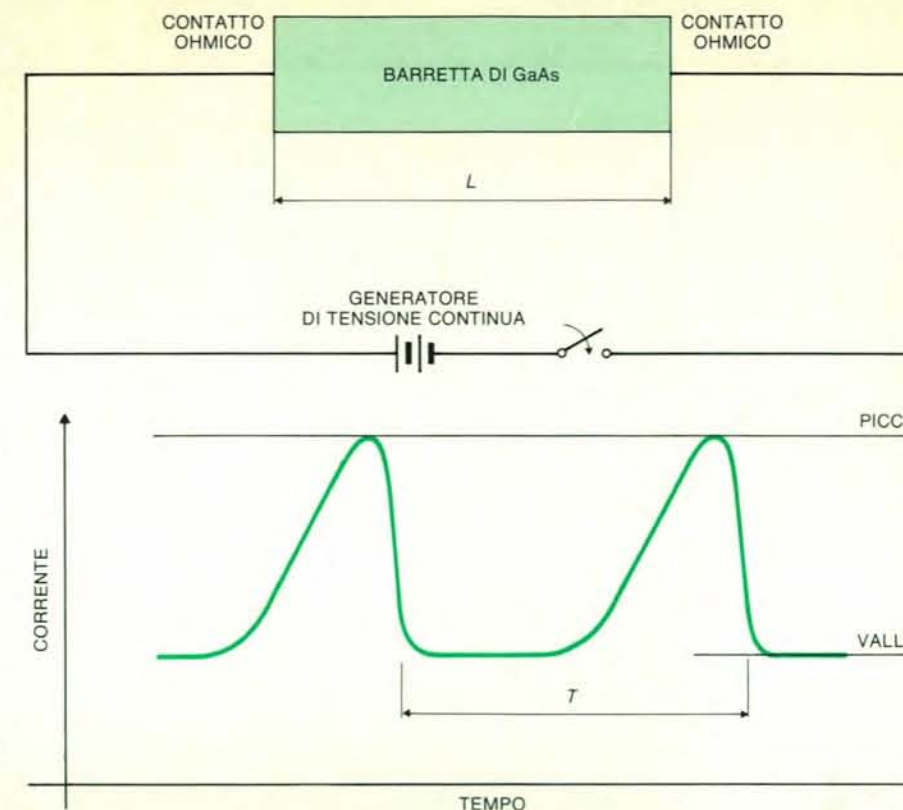
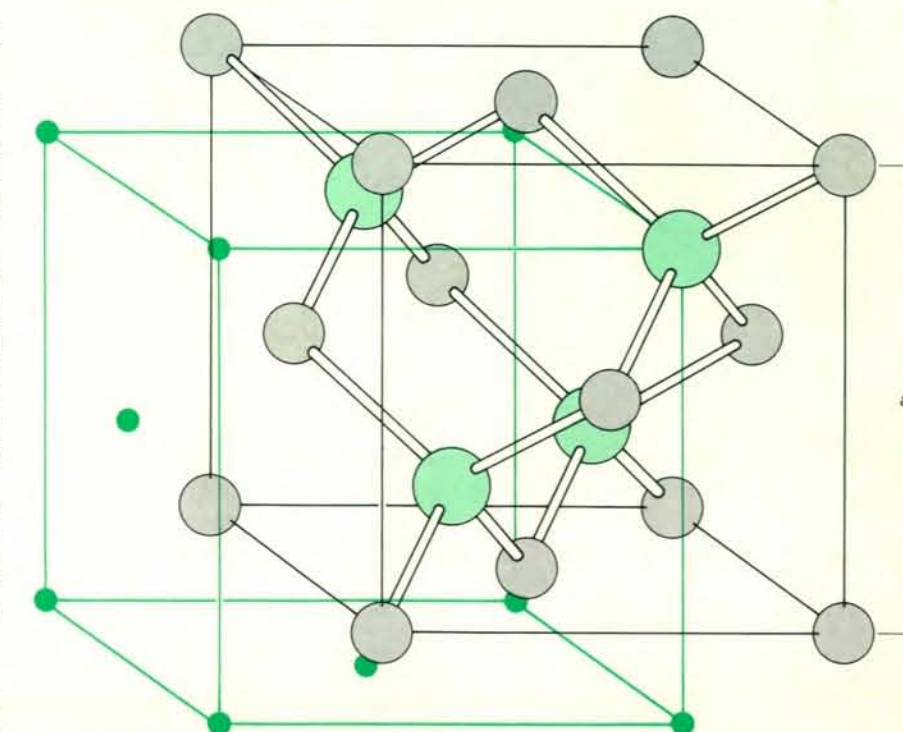
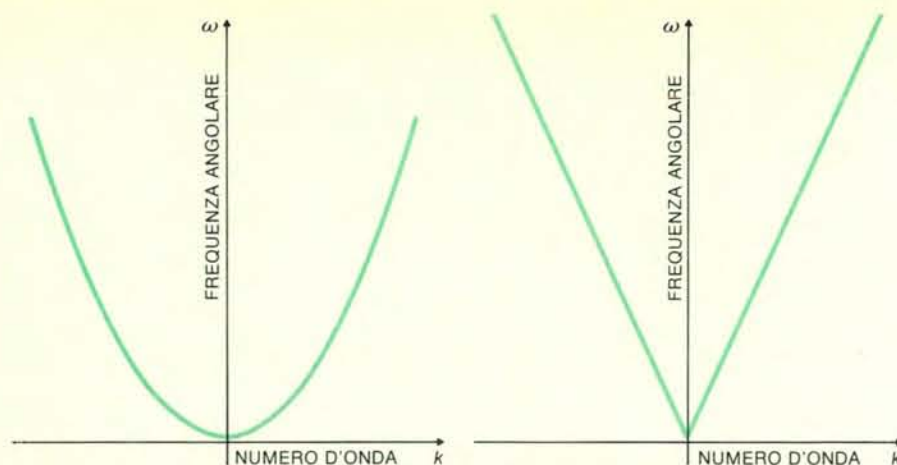
La situazione si modifica completamente quando l'elettrone è costretto a muoversi in un potenziale periodico come quello di un cristallo. Nel semplice caso unidimensionale infinito si giunge alla conclusione che non tutti i valori dell'energia sono permessi. I valori possibili sono raggruppati in bande separate tra loro da intervalli *(gap)* proibiti. Inoltre, all'interno di ogni banda l'energia è una funzione periodica del numero d'onda e l'ampiezza delle bande permesse aumenta al crescere dell'energia. Il problema è del tutto analogo alla propagazione di un'onda elastica o di un'onda elettromagnetica in una struttura periodica; in entrambe le circostanze i diagrammi di dispersione presentano bande di arresto in diversi intervalli di frequenza.

Il confronto tra le curve di dispersione della particella libera e nel cristallo unidimensionale mette in luce che, nel secondo caso, l'energia non è proporzionale al quadrato della quantità di moto. Tale proporzionalità può essere però ricreata in sede teorica adottando l'approssimazione della *massa effettiva,* ovvero supponendo che l'elettrone si muova nel potenziale periodico come se fosse libero ma dotato di una massa m^* proporzionale in ogni punto al raggio di curvatura - in quel punto - della curva di dispersione. Pertanto, dal diagramma a bande appare che, oltre al fatto di avere una massa variabile all'interno di ogni banda, gli elettroni dotati di minore energia (i più legati ai «pozzi» di potenziale del cristallo unidimensionale) sono rappresentati da un ramo a piccola curvatura, cioè a grande raggio; la loro notevole massa effettiva fa sì che l'eventuale applicazione di un campo elettrico esterno non li influenzi apprezzabilmente, almeno per valori non elevatissimi dell'intensità.

Il modello del cristallo infinito unidimensionale può essere applicato ai cristalli reali. Naturalmente, bisogna usare una certa cautela e tener conto di limitazioni inevitabili; tra queste, le più importanti in vista delle applicazioni che ne faremo in seguito sono l'anisotropia del cristallo reale e le sue dimensioni finite.

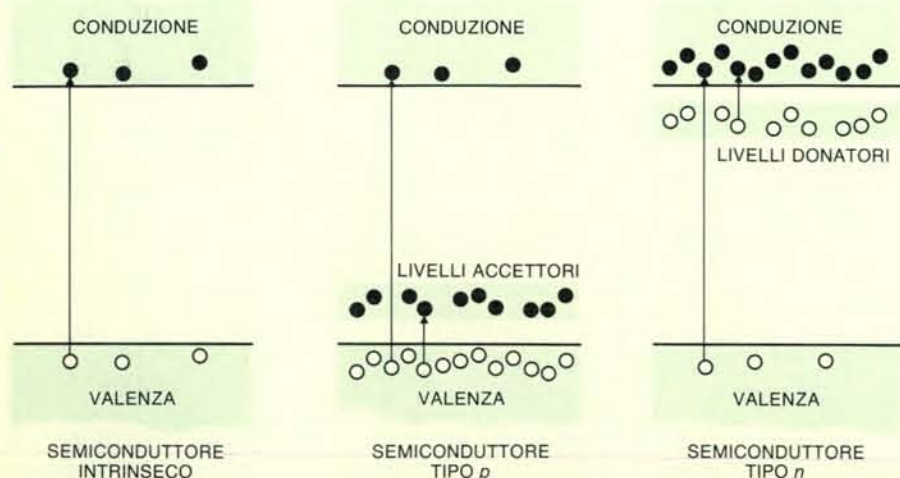
Il primo problema nasce immediatamente considerando che gli elettroni possono muoversi in tre dimensioni. Il numero d'onda k diviene quindi una entità vettorale **k**, dovendo precisare una direzione nello spazio oltre che una grandezza. Le componenti del vettore d'onda vengono riferite a tre direzioni, arbitrarie purché non complanari, all'interno del cristallo: solitamente vengono scelti i tre spigoli della cella elementare. In tali condizioni, il diagramma ω **(k)** dovrebbe essere disegnato in quattro dimensioni poiché è necessario associare un numero (la frequenza o, equivalentemente, l'energia) a ogni punto dello spazio dei vettori d'onda **k**. Per semplificare, si può considerare di volta in volta la direzione che interessa e tracciare la curva di dispersione per quella direzione. Naturalmente, poiché la periodicità del cristallo varia in funzione della direzione prescelta (anisotropia), i diagrammi ω**(k)** varieranno di conseguenza.

Analizzando a fondo l'altro aspetto, cioè il fatto che un cristallo reale non è infinito, si giunge alla conclusione che **k** non può assumere qualsiasi valore (in modulo, o lunghezza, del vettore); sono possibili, all'interno di ogni banda e per ogni direzione, un numero finito di *stati*, ovvero un numero finito di valori per il numero d'onda ai quali ne corrispondono altrettanti per l'energia. Globalmente, in ogni banda il numero di stati energetici permessi è pari a quello delle unità di forma del cristallo. Per il principio di esclusione di Pauli, ogni stato può essere occupato al più da due elettroni, purché abbiano spin opposti (lo spin è la rotazione intorno al proprio asse dell'elettrone schematizzato come una sferetta rigida). In altre parole, all'interno del cristallo non vi possono essere più di due elettroni viaggianti con la stessa velocità nella medesima direzione.

Mediante la teoria delle bande nei solidi è possibile comprendere il diverso comportamento dei vari materiali in risposta all'azione di un campo elettrico. A questo proposito, è sufficiente prendere in considerazione la banda più

In alto, nella figura, è presentato lo schema di principio del circuito impiegato da J. B. Gunn nei suoi esperimenti al Thomas J. Watson Research Center della International Business Machines Corporation, nel 1963, da cui ha avuto inizio lo sviluppo del diodo Gunn. Nella parte inferiore della figura è presentata la forma d'onda della corrente ottenuta quando la tensione *V* è sufficiente a generare campi elettrici dell'ordine di migliaia di volt per centimetro. Quando il campo elettrico è stato portato al di sopra del valore di soglia e hanno avuto inizio le oscillazioni, i valori di picco e di valle della corrente rimangono praticamente costanti al crescere della tensione applicata. Il periodo delle oscillazioni risulta invece proporzionale alla lunghezza del campione; nel corso dei suoi esperimenti Gunn utilizzò barrette di lunghezza variabile tra 20 e 200 micrometri.

Struttura cristallina del GaAs. Il reticolo è cubico a facce centrate e l'unità di forma è costituita da un atomo di arsenico e uno di gallio; la distanza tra essi è un quarto della diagonale di corpo del cubo di lato *a*. Assumendo tale cubo come cella elementare si hanno quattro unità di forma per cella.

Nel caso di un elettrone «libero» l'onda a esso associata è caratterizzata da una frequenza angolare e da un numero d'onda, direttamente proporzionali rispettivamente all'energia e alla quantità di moto della particella. La relazione tra queste due caratteristiche viene definita *curva di dispersione*. A sinistra nell'illustrazione è rappresentata la curva di dispersione dell'elettrone libero: poiché l'energia è direttamente proporzionale al quadrato del momento lineare, la curva di dispersione è una parabola. A destra è tracciata l'analoga curva per un'onda elastica in un mezzo continuo.

alta, nella scala delle energie, avente gli stati completamente occupati da elettroni e l'adiacente superiore, vuota o solo parzialmente riempita nei suoi livelli più bassi; esse vengono di solito indicate con i nomi di *banda di valenza* e *banda di conduzione*, rispettivamente. Gli elettroni delle bande inferiori non vengono considerati in quanto, come s'era accennato precedentemente, non partecipano alla conduzione elettrica a causa della loro grande massa effettiva.

Quando si applica un campo elettrico al cristallo, può instaurarsi un flusso di cariche, cioè una corrente, solo se tali portatori (elettroni) sono in grado di acquistare l'energia fornita loro dal campo esterno. Affinché ciò sia possibile, gli elettroni devono poter «risalire» il diagramma a bande e trovare degli stati energetici vuoti ove collocarsi.

Un isolante è caratterizzato dal fatto di avere la banda di valenza completamente saturata e quella di conduzione del tutto vuota, separate da un intervallo proibito piuttosto grande (per esempio, nel diamante a temperatura ambiente l'ampiezza dell'intervallo proibito è di circa 5,3 eV). Sottoponendo all'azione di un campo elettrico un simile materiale non si ha alcuna corrente perché gli elettroni non trovano superiormente livelli energetici liberi, a meno di «saltare» l'intervallo proibito e portarsi in banda di conduzione. Data l'ampiezza dell'intervallo proibito, ciò richiederebbe l'impiego di campi di notevole intensità.

Nel semiconduttore intrinseco non vi sono livelli energetici permessi all'interno dell'intervallo proibito. La concentrazione di elettroni in banda di conduzione è uguale a quella delle buche in banda di valenza. I semiconduttori di tipo *p* e *n*, drogati con atomi di impurità di elementi appartenenti al III e V gruppo rispettivamente, presentano al contrario una fascia di energie permesse all'interno dell'intervallo proibito. La conduzione avviene prevalentemente per mezzo delle buche nella banda di valenza (tipo *p*) o degli elettroni nella banda di conduzione (tipo *n*).

Quando la banda di conduzione è parzialmente occupata, il materiale si comporta come un buon conduttore di elettricità per le ragioni opposte al caso precedente: gli elettroni possono acquistare energia muovendosi verso gli stati liberi superiori della banda di conduzione.

In un semiconduttore le due bande in questione sono strutturate come nell'isolante, salvo l'ampiezza dell'intervallo proibito che, in questo caso, è relativamente piccolo. Per limitarsi ai materiali più noti, si hanno valori che vanno da 0,67 eV per il germanio a 1,14 eV per il silicio fino a 1,4 eV per l'arseniuro di gallio.

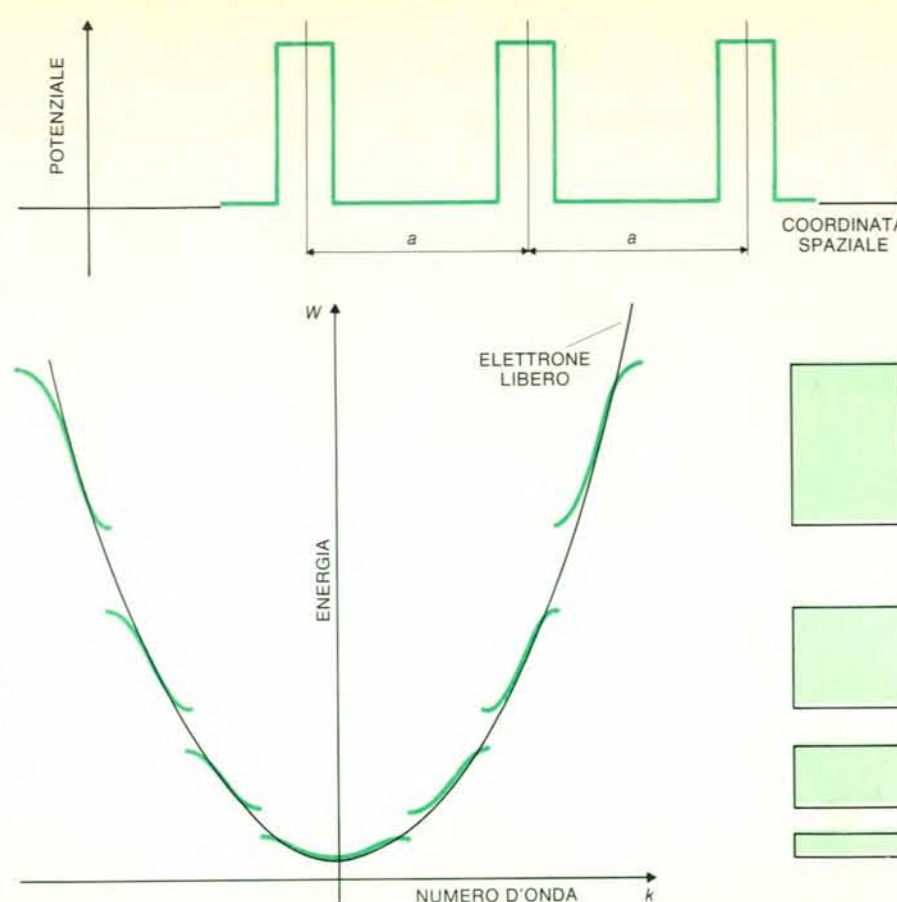
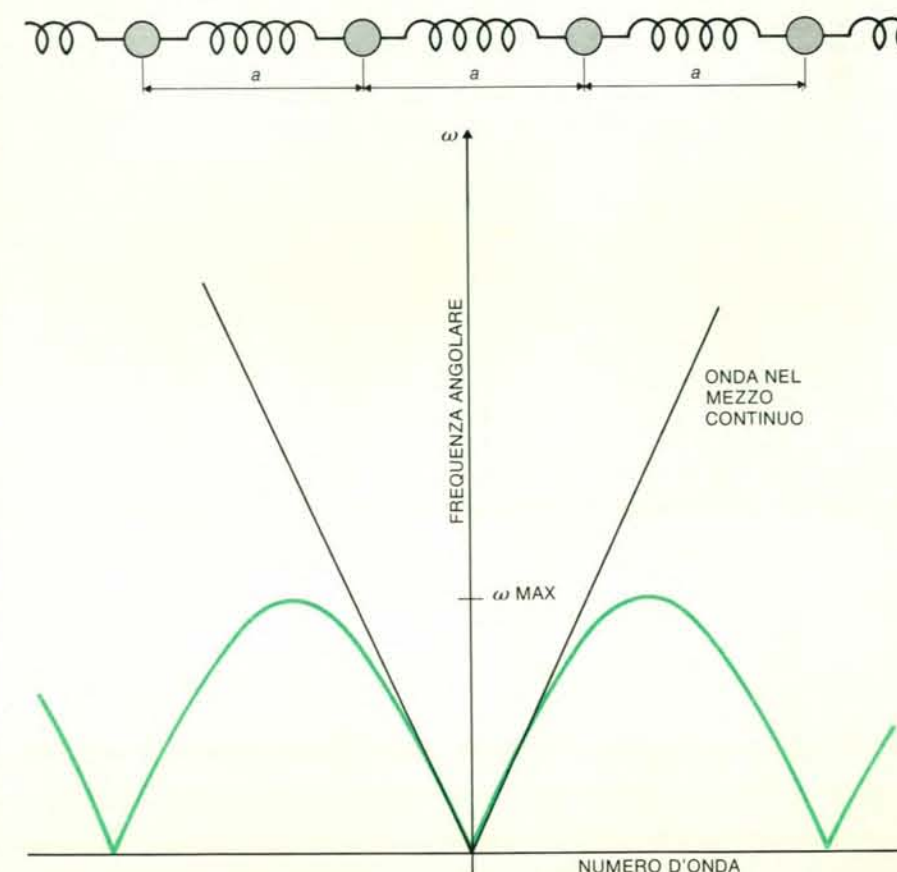
Allo zero assoluto, non essendovi apporto di energia di alcun genere dall'esterno, la banda di valenza è completamente satura e quella di conduzione vuota; dal punto di vista elettrico il materiale è un isolante. All'aumentare della temperatura, un numero sempre maggiore di elettroni acquista energia sufficiente per passare, superando l'intervallo proibito, in banda di conduzione. Si vanno perciò rendendo disponibili un numero via via maggiore di portatori di carica in banda di conduzione e un uguale numero di «buchi», ossia di stati energetici liberi, in banda di valenza. Conseguentemente, il materiale diviene sempre più conduttore.

Il più delle volte però questa conducibilità intrinseca non viene direttamente utilizzata in pratica. Infatti, i portatori intrinseci sono una piccola quantità rispetto a quelli generati da impurezze introdotte appositamente nel semiconduttore (come si usa dire, «drogandolo») per esaltare il fenomeno della conduzione. Tali impurità sono costituite da atomi di elementi appartenenti al III gruppo - come il boro o l'indio - oppure al V - come l'arsenico o il fosforo - in grado, rispettivamente, di catturare un elettrone o di cederlo al cristallo creando così uno squilibrio rispetto al caso del semiconduttore puro. Nello schema a bande le impurità hanno l'effetto di introdurre dei livelli energetici permessi all'interno dell'intervallo proibito. Nel caso si tratti di atomi del III gruppo *(accettori)* diviene disponibile una piccola fascia di livelli, completamente vuoti allo zero assoluto, immediatamente sopra la banda di valenza; a temperatura ambiente, essa sarà molto più svuotata che in un semiconduttore intrinseco poiché, oltre agli elettroni andati a occupare gli stati più bassi della banda di conduzione, una certa quantità - molto maggiore della precedente - avrà saturati i livelli accettori. Si usa dire che il semiconduttore è *estrinseco* di tipo *p*, sottolineando così il fatto che i portatori mobili di carica sono i buchi positivi della banda di valenza (causati dalla mancanza di elettroni).

Analogamente, le impurezze del V gruppo vengono dette *donatori* perché a temperatura ambiente hanno ceduto alla banda di conduzione il loro elettrone in eccesso. Ciò è possibile in quanto i donatori introducono una fascia di energie permesse appena sotto la banda di conduzione. Un semiconduttore drogato con donatori viene detto di tipo *n* per indicare che i portatori maggioritari di carica (cioè i più numerosi) sono elettroni in banda di conduzione. Impiegando entrambi i tipi di semiconduttori estrinseci furono realizzate dapprima le giunzioni *p-n* rettificatrici e in seguito il transistore a giunzione (mediante due giunzioni *p-n*).

A differenza di questi ultimi, il diodo Gunn non è costituito da alcuna giunzione essendo, come s'è detto inizialmente, una barretta di composto del III - V gruppo (il più usato è il GaAs) drogato con donatori. Le ragioni del suo comportamento devono perciò ricondursi esclusivamente a effetti di volume, ovvero del solo semiconduttore, piuttosto che a proprietà legate alla giunzione *p-n*. Inoltre, mentre nella fisica del transistore gli elettroni che partecipano alla conduzione sono resi disponibili dalla energia termica dell'ambiente, dell'ordine di qualche decina di meV (millesimi di elettronvolt), l'effetto Gunn si manifesta esclusivamente sotto l'azione di campi elettrici elevati: gli elettroni coinvolti nel fenomeno devono essere dunque altamente energetici. La teoria RWH (Ridley, Watkins, Hilsum) prevede infatti energie dell'ordine delle centinaia di meV, equivalenti a quelle che si otterrebbero a temperature di qualche migliaio di gradi; per questa ragione, gli elettroni di conduzione nel diodo Gunn vengono detti «caldi».

La spiegazione del fenomeno si basa sull'aspetto particolare del diagramma a bande dell'arseniuro di gallio (o di altri composti idonei, come il fosfuro di indio InP). Nella illustrazione della pagina seguente è mostrata schematicamente la sezione bidimensionale del grafico nelle direzioni [100] e [111], ovvero lungo uno spigolo e la diagonale di corpo del cubo a facce centrate, rispettivamente. Come appare dalla figura, in entrambe le direzioni il profilo della banda di conduzione è caratterizzato da un minimo principale, distante 1,4 eV dalla cima della banda di valenza e da un minimo secondario. La separazione, su una scala di energie, tra la valle principale e le secondarie dipende dalla direzione prescelta all'interno del cristallo: il valore minimo in assoluto di 0,36 eV lo si trova procedendo parallelamente agli spigoli del cubo. Comunque, per le considerazioni che seguiranno è sufficiente ragionare su un semplice schema a due valli della banda di conduzione.

A temperatura ambiente i donatori avranno ceduti completamente i loro elettroni che, dovendosi collocare sul fondo della banda di conduzione data la modesta entità dell'energia termica, riempiranno gli stati più profondi della valle principale. La differenza di energia tra i due minimi è abbastanza grande da garantire che la valle satellite è completamente vuota. Applicando un campo elettrico esterno di intensità via via crescente, si avrà come conseguenza un incremento della velocità media di trascinamento *(drift)* degli elettroni nella direzione del campo, secondo un coefficiente di proporzionalità che viene detto «mobi-

Nella figura sopra è mostrato l'andamento del potenziale periodico secondo il modello unidimensionale del cristallo di Kronig-Penney. Il diagramma di dispersione in colore mostra le bande permesse a cui corrispondono i valori possibili per l'energia; a destra è lo schema a bande semplificato. La parabola tratteggiata è la *W(k)* per l'elettrone libero. Sotto, la curva di dispersione per un'onda elastica propagantesi nel mezzo periodico costituito da masse interagenti tramite molle.

lità» dei portatori. Questa situazione permane fino a che il campo non supera un certo valore di soglia, in corrispondenza del quale gli elettroni ricevono energia sufficiente per passare direttamente nella valle secondaria. Aumentando ancora il campo, si giunge a trasferire tutti gli elettroni della valle principale in quella satellite, occupando gli stati posti sul fondo di quest'ultima. Nel corso di questo trasferimento accade però un fatto molto importante: gli elettroni acquistano una massa effettiva m^* più grande, come appare immediatamente osservando le differenti curvature delle due valli. Nel caso del GaAs, indicando con m_0 la massa dell'elettrone libero, si ha $m^* = 0{,}072\ m_0$ sul minimo principale e $m^* = 1{,}2\ m_0$ sul minimo secondario. Conseguentemente, nel GaAs gli elettroni della banda di conduzione passano, all'aumentare del campo, da una mobilità di 8000 cm^2/volt per secondo al valore di 150. Il comportamento del materiale nella fase di transizione è di importanza fondamentale; infatti, se un incremento del campo provoca un sufficiente «riscaldamento» dei portatori, ci si può trovare in presenza di una mobilità differenziale negativa. In altre parole, un incremento del campo provoca contemporaneamente un'accelerazione delle cariche e un loro «appesantimento»; se il secondo effetto prevale, la velocità di trascinamento diminuisce all'aumentare del campo. In pratica, ciò è quanto accade in numerosi composti del III-V gruppo e nell'arseniuro di gallio in particolare: nella illustrazione in alto nella pagina a fronte è mostrato, indicativamente, l'andamento della curva $v(E)$ (velocità di trascinamento/campo elettrico) per tali materiali.

Le instabilità osservate da Gunn furono appunto causate dalla presenza di questo tratto di caratteristica a mobilità differenziale negativa. A prima vista, l'andamento della curva $v(E)$ sembra suggerire come ovvio un aspetto analogo per la caratteristica statica tensione/corrente del dispositivo, assumendo che la corrente sia proporzionale alla velocità e la tensione al campo elettrico. Se così fosse, si avrebbe un tratto di curva in cui la corrente diminuisce all'aumentare della tensione; si renderebbe in tal modo disponibile una conduttanza differenziale (cioè la pendenza della caratteristica tensione/corrente) *negativa*, in grado di funzionare come un generatore in regime alternativo, alla quale associare il meccanismo delle oscillazioni di Gunn. Il comportamento del dispositivo sarebbe perciò simile a quello di una particolare giunzione *p-n* nota come «diodo tunnel» (realizzata nel 1958 da L. Esaki). In realtà, la proporzionalità tra il campo elettrico e la tensione applicata sussiste solo nel caso che il primo sia uniforme, il che è abbastanza difficile, se non impossibile, da ottenere in pratica. Inoltre, lo stesso Shockley (l'inventore del transistor a giunzione) dimostrò nel 1954 che una diminuzione della velocità di trascinamento in seguito a un incremento del campo non porta, in generale, a un decremento della corrente con l'aumentare della tensione applicata.

L'effetto Gunn è invece legato al fatto che un semiconduttore con mobilità differenziale negativa è intrinsecamente instabile, nel senso che preciseremo immediatamente. Si supponga di sottoporre la barretta di GaAs a un campo superiore al valore di soglia e che, a causa di un difetto nel cristallo, una disuniformità nel drogaggio o, al limite, una semplice fluttuazione statistica, si abbia in un certo punto del campione un accumulo di cariche. Se il semiconduttore fosse normale, cioè con mobilità differenziale positiva, esso tenderebbe a smaltire l'eccesso di carica con una costante di tempo (tempo di rilassamento) dipendente dalle sue caratteristiche e in primo luogo dalla concentrazione dei portatori. Ricorrendo a una analogia, si può paragonare questo comportamento a quello di un liquido che, sottoposto a un disturbo, tende a ristabilire l'orizzontalità della superficie libera in un tempo più o meno lungo a seconda delle sue proprietà (come, per esempio, la sua viscosità).

In un semiconduttore con mobilità differenziale negativa avviene esattamente il contrario: un accumulo di carica spaziale in una regione produce un incremento del campo elettrico locale al quale corrisponde un rallentamento dei portatori di carica. La conseguenza immediata di questo fatto è un ulteriore accumulo delle cariche provenienti dal catodo (polo negativo) mentre in direzione dell'anodo (polo positivo) si crea un vuoto di cariche; ciò produce uno strato di dipolo che tende a rinforzare ancor più il campo elettrico. Pertanto, il fenomeno si esalta spontaneamente fino al raggiungimento di una situazione di equilibrio nella quale un dominio di dipolo «maturo» - cioè avente una configurazione stabile - viaggia verso l'anodo. L'equilibrio è possibile in quanto ai capi del campione viene mantenuta una differenza di potenziale costante; conseguentemente, via via che il campo elettrico aumenta all'interno del dominio si deve avere una sua diminuzione compensativa all'esterno. Applicando questa considerazione alla caratteristica $v(E)$, come è stato visualizzato nell'illustrazione in alto nella pagina a fronte, il punto *D* rappresentativo del campo dipolare si sposterà verso destra mentre *C*, associato al campo esterno, muoverà verso sinistra. La configurazione di equilibrio viene raggiunta quando *D* e *C* si stabilizzano sul medesimo valore della velocità in modo da determinare univocamente la corrente nel cristallo. In questa situazione, il dominio di alto campo viaggia verso l'anodo mantenendo un assetto stabile. Tra l'altro, occorre notare che la diminuzione fin sotto la soglia del campo all'esterno del dominio impedisce la formazione di nuovi strati di dipolo.

Schema a bande per l'arseniuro di gallio. Le parti in colore si riferiscono alla direzione [111], mentre quelle in grigio alla direzione [100]. In entrambe le direzioni il fondo della banda di conduzione è caratterizzato da una coppia di minimi a differente curvatura separati da qualche decimo di eV. Il funzionamento del diodo Gunn è in rapporto al trasferimento degli elettroni dal minimo inferiore a quello superiore.

Per chiarire meglio il processo descritto, possiamo ricorrere nuovamente a una immagine comune associando al flusso delle cariche nel campione il movimento di una colonna di automezzi militari. Se, per una ragione qualsiasi, uno dei veicoli centrali rallenta, alcuni mezzi immediatamente seguenti tenderanno ad avvicinarlo «impacchettandosi», mentre quelli che lo precedevano, proseguendo indisturbati, creeranno una zona di svuotamento davanti al pacchetto. Però, poiché una delle prerogative della colonna militare è di arrivare completa a destinazione e non in due tronconi, gli automezzi che precedono il pacchetto, resisi conto della situazione, diminuiranno la loro velocità in modo da non guadagnare terreno ulteriormente. Analogamente, quelli che seguono la zona di accumulazione rallenteranno per mantenere le distanze iniziali. In definitiva, la velocità dell'intera colonna sarà diminuita e il «dipolo di automezzi», cioè l'intera zona comprendente l'accumulazione e lo svuotamento, si muoverà con questa velocità - mantenendosi stabile - all'interno della colonna stessa.

Una volta raggiunto l'anodo, il dominio «collassa» rapidamente provocando un impulso di corrente nel circuito esterno. D'altro canto, non appena la carica spaziale scompare il campo elettrico all'interno del campione ritorna al valore iniziale e ha inizio un nuovo ciclo di formazione di un dominio. Nonostante vi possano essere più punti di discontinuità in grado di funzionare come centri di enucleazione del dominio, il principale è il catodo (dove si hanno le maggiori disuniformità del drogaggio).

La teoria dei domini di dipolo rende conto perfettamente delle osservazioni sperimentali di Gunn anche dal punto di vista del periodo delle oscillazioni. Infatti, l'intervallo di tempo tra un impulso di corrente e il successivo dipende esclusivamente dal tempo di transito del dominio nel campione, ovvero dal rapporto L/v tra la lunghezza della barretta e la velocità di trascinamento all'equilibrio; poiché quest'ultima è una caratteristica del semiconduttore, ne viene che il tempo di transito dipende, a parità di altre condizioni, dalla lunghezza del campione.

Lo stesso Gunn eseguì successivamente delle esperienze molto raffinate mediante le quali, usando delle sonde accoppiate capacitivamente alla barretta, riuscì a evidenziare la presenza dei domini ad alto campo e il loro movimento nel cristallo. Questo risultato sperimentale si rivelò utilissimo per l'approfondimento del fenomeno e stimolò una quantità enorme di ricerche e studi in questa direzione. Però, esso contribuì involontariamente a creare, almeno inizialmente, la convinzione errata che gli oscillatori a microonde - realizzati nel frattempo con simili dispositivi - funzionassero tutti secondo il principio descritto.

In realtà, il *modo Gunn* non viene attualmente impiegato nella attuazione pratica di generatori ad alta frequenza perché, dipendendo esclusivamente dal tempo di transito, la frequenza delle oscillazioni non può essere variata agendo sul circuito esterno.

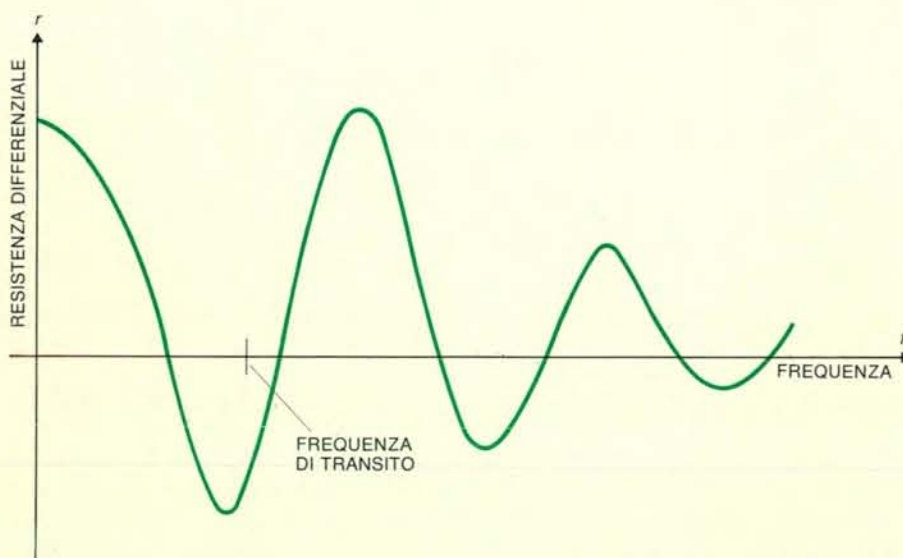
La situazione è molto diversa quando si inserisce il diodo Gunn in un circuito risonante a basse perdite, come può essere una cavità a microonde. In tal caso, l'ampiezza della tensione ad alta frequenza che si sviluppa ai capi del dispositivo può diventare dello stesso ordine di grandezza della tensione continua di polarizzazione (superiore al valore di soglia): in particolare, può accadere che la tensione totale scenda al disotto del valore di soglia per una frazione del ciclo della microonda.

Il risultato di questa interazione tra circuito e diodo è che si creano due altre possibilità di funzionamento. Nel cosiddetto *modo a dominio spento* vengono sostenute oscillazioni con periodo minore del tempo di transito; ciò è possibile in quanto il dominio maturo in viaggio verso l'anodo viene spento prima di raggiungerlo per il decadere della tensione al disotto della soglia. Al contrario, nel *modo a dominio ritardato* il periodo della microonda è maggiore del tempo di transito. In tal caso, lo strato di dipolo attraversa il campione in tutta la sua lunghezza ma la sua formazione viene ritardata di un tempo pari alla frazione di ciclo durante la quale il campo è al disotto della soglia.

In entrambi i casi le oscillazioni avvengono a una frequenza dipendente dalle caratteristiche del circuito piuttosto che da quelle del dispositivo, consentendo

Andamento caratteristico della velocità di trascinamento in funzione della intensità del campo elettrico per l'arseniuro di gallio. Il tratto a mobilità differenziale negativa raccorda le due rette che rappresentano l'alta e la bassa mobilità nelle valli principale e secondaria, rispettivamente.

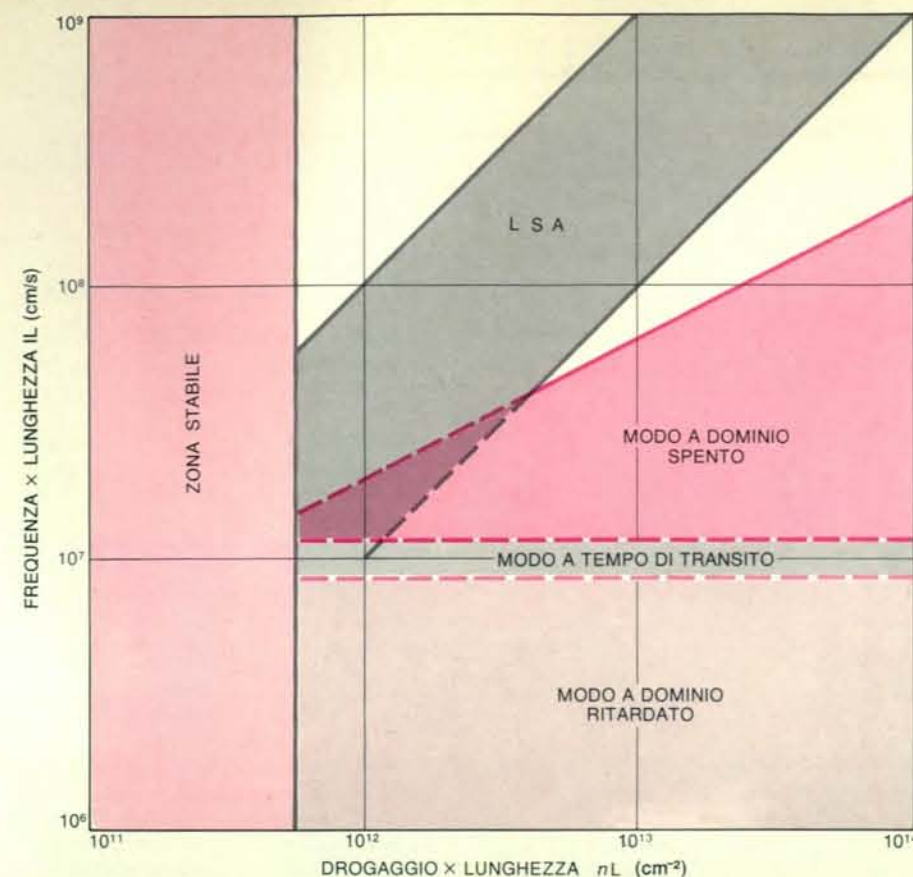
Campo elettrico e densità dei portatori all'interno della barretta di arseniuro di gallio. Sono mostrati il dominio ad alto campo e il corrispondente strato dipolare viaggianti verso l'anodo.

Quando un diodo Gunn viene inserito in un circuito risonante a basse perdite, come la cavità coassiale mostrata in alto, la tensione alternata ad alta frequenza che si sviluppa ai suoi capi può portare il campo sotto il valore di soglia per una frazione di ciclo. Nel grafico è visualizzato qualitativamente il fenomeno: nell'intervallo di tempo *(T_1, T_2)* il campo è minore del valore di soglia.

Andamento tipico della resistenza differenziale in funzione della frequenza per un campione «sottocritico», caratterizzato cioè da un valore nL minore di $7,6 \times 10^{11}$ cm^{-2}. Nell'intorno della frequenza di transito e delle sue armoniche si hanno delle bande a resistenza differenziale negativa.

così la realizzazione di generatori nei quali la frequenza può essere variata agendo su un elemento circuitale.

Nel descrivere i vari modi di funzionamento si è fatta implicitamente l'ipotesi che il tempo di maturazione del dominio fosse trascurabile rispetto agli altri tempi in gioco. In realtà, tale condizione è soddisfatta quando il rapporto tra la concentrazione n e la frequenza f supera il valore di 5×10^4 sec/cm^3. L'importanza del drogaggio in queste considerazioni deriva dal fatto che il tempo di rilassamento del semiconduttore, ovvero la costante di tempo nel processo di formazione dei domini, è direttamente proporzionale a questa grandezza. Tra l'altro, essa entra in un secondo parametro, il prodotto nL (essendo L la lunghezza della barretta), il quale ha un ruolo fondamentale nel separare le classi dei modi di funzionamento basati sulla presenza dei domini da altre classi di modi di funzionamento nelle quali non vi sono accumuli di carica spaziale nel campione.

La criticità di questo parametro per quel che concerne la stabilità del campione è intuitiva; infatti, affinché si possa formare uno strato di dipolo maturo è indispensabile che il processo di maturazione duri meno del tempo di transito, il che implica una certa compatibilità tra la lunghezza L della barretta e la concentrazione n dei portatori. Da questo punto di vista, un campione corto e fortemente drogato può essere equivalente a uno più lungo ma drogato in misura minore. Nel primo caso, il dominio maturerà per via della piccola costante di tempo con cui avviene l'accumulazione; nel secondo il processo è più lento ma il cammino disponibile è più grande. Un'analisi approfondita, confortata dai risultati sperimentali, ha permesso di stabilire che per valori di nL minori di $7,6 \times 10^{11}$ cm^{-2} (nel caso di GaAs) ogni campione è stabile, nel senso che non può sostenere la formazione e la propagazione di domini di carica spaziale.

Nonostante ciò, l'impedenza di tali dispositivi presenta degli intervalli di frequenza, centrati intorno alla frequenza di transito e alle sue armoniche, nei quali la resistenza è negativa. Ne viene che, sebbene sia stabile in bassa frequenza, la barretta può essere collocata in una cavità a microonde nella quale, regolando opportunamente la frequenza di risonanza e il livello di impedenza, potranno avvenire le oscillazioni (pur non essendovi formazione di domini di carica spaziale). In questo modo di funzionamento il diodo Gunn si comporta similmente ad altri dispositivi a resistenza negativa - come il diodo tunnel - limitatamente però alle bande di frequenza di cui s'è detto in precedenza. Per completare il quadro dei vari modi di funzionamento è necessario descrivere quello che, per molti aspetti, dovrebbe essere considerato come il fondamentale. Esso viene indicato con il nome di LSA (Limited Space charge Accumulation, accumulazione limitata di carica spaziale) e fu scoperto da J. A. Copeland nel 1966 durante una simulazione al calcolatore.

Carta dei modi di funzionamento del diodo Gunn. Quando il prodotto lunghezza-drogaggio è inferiore a ~ $7,6 \times 10^{11}$ cm^{-2} il campione è stabile. Per valori superiori si hanno le varie modalità di oscillazione in funzione del prodotto frequenza-lunghezza. Il punto di lavoro dipende dal circuito a microonde collegato al dispositivo; nella zona di sovrapposizione ($fL \simeq 10^7$, $nL \simeq 10^{13}$) la tensione di polarizzazione del diodo è fondamentale per la determinazione del modo di funzionamento.

In questo modo di funzionamento, il campione si avvicina al comportamento ideale - il più semplice concettualmente - di un semiconduttore drogato uniformemente, con contatti ohmici paralleli e privo di qualunque carica spaziale interna. In tale circostanza il campo elettrico verrebbe a essere uniforme e proporzionale alla tensione applicata; la caratteristica statica tensione/corrente del dispositivo avrebbe lo stesso andamento della curva $v(E)$ presentando così un tratto a resistenza differenziale negativa che, come nel caso precedente (ma senza limitazioni di frequenza), potrebbe essere utilizzata in un circuito oscillatorio. In realtà, tale uniformità del campo viene ottenuta quando la radiofrequenza ha un'ampiezza sufficiente per portare, a ogni ciclo, il campione sotto la soglia e una frequenza elevata al punto che il dominio, in via di enucleazione al catodo, non riesce a staccarsi per viaggiare verso l'anodo prima che il campo scenda nuovamente sotto la soglia, spegnendolo. Conseguentemente, la zona esterna al dominio, nella quale il campo si abbassa durante l'accumulazione, è costituita in pratica dall'intero campione. Un calcolo dettagliato dimostra che tale modo di funzionamento può essere sostenuto quando il rapporto n/f è compreso all'incirca tra 10^4 e 10^5 secondi per cm^3.

Per concludere, possiamo citare alcuni dati pratici sugli oscillatori a diodo Gunn e accennare brevemente agli sviluppi futuri che si prevedono per tali dispositivi. A questo scopo, occorre fare una distinzione fra componenti realizzati con GaAs, tecnologicamente maturi, e quelli a InP tuttora in fase di sviluppo. Mediante la prima famiglia vengono attualmente generate potenze comprese tra 1 watt (a 8 GHz) e 0,05 watt (a 50 GHz) con efficienze tipiche del 3-5 per cento (massime del 10-15 per cento). I dispositivi a InP sembrano invece destinati a diventare eccellenti generatori nel campo delle onde millimetriche, con efficienze almeno doppie dei precedenti. Le realizzazioni più recenti hanno raggiunto potenze comprese tra 0,2 watt (a 50 GHz) e 0,1 watt (a 70 GHz).

Per quanto riguarda gli sviluppi futuri, si prevede che il diodo Gunn, sia GaAs che InP, sarà ancora per lungo tempo il generatore a basso rumore più economico e affidabile. Per le applicazioni nelle quali non è richiesta la bassa rumorosità esso sarà soppiantato da altri due dispositivi allo stato solido: il FET (Field Effect Transistor, transistore a effetto di campo) al di sotto di 10 GHz e il diodo IMPATT, una giunzione *p-n* che lavora nella regione di ionizzazione a valanga, nella regione superiore dello spettro.

Il parsec centrale della Galassia

Le osservazioni nel radio e nell'infrarosso indicano che il centro della nostra galassia ospita un oggetto supercompatto, forse un buco nero massiccio, immerso in un denso vortice di stelle, gas e polvere

di Thomas R. Geballe

All'inizio degli anni trenta Karl Jansky, un giovane fisico impiegato presso i Bell Telephone Laboratories nel New Jersey, costruì il primo radiotelescopio perché lo aiutasse nell'identificazione di sorgenti di interferenze che potessero disturbare le comunicazioni radiotelefoniche transatlantiche. Una delle sorgenti principali di disturbo scoperte da Jansky era un «sibilo stazionario», come egli lo descrisse, che proveniva dal cielo e raggiungeva la massima intensità in una direzione molto a sud del suo strumento. Col passare del tempo osservò che la sorgente di questa interferenza seguiva il moto delle stelle fisse e pertanto aveva origine al di fuori del sistema solare. La direzione del picco di intensità, verso la costellazione del Sagittario, coincideva con quella che già gli astronomi consideravano come la direzione del centro della nostra galassia.

Le osservazioni di Jansky non costituirono soltanto le prime osservazioni radioastronomiche, ma anche la prima rivelazione di una qualsiasi onda elettromagnetica proveniente dal nucleo della Galassia, che è nascosto da nubi di polvere interstellare. In seguito, coi progressi in elettronica durante la seconda guerra mondiale, la radioastronomia acquistò credito come progetto scientifico dotato di tutte le migliori qualità. Negli anni sessanta e settanta la tecnologia infrarossa, che era rimasta indietro rispetto alla controparte radio, cominciò a fornire rivelatori e spettrometri di buona sensibilità che potevano essere usati con grandi telescopi ottici. Con la loro capacità di attraversare le nubi di polvere interposte, la radioastronomia e l'astronomia infrarossa hanno contribuito a formare la maggior parte delle conoscenze attuali sulla regione centrale della Galassia. Negli ultimi anni continui progressi in entrambi i campi hanno consentito agli astronomi di iniziare a esplorare in dettaglio la regione più interna di elevatissima densità.

I nuclei galattici sono molto densi per effetto soltanto della gravitazione. Si pensa che le galassie nell'universo si siano formate a partire da disuniformità presenti nel gas emesso dal «big bang», in cui nacque l'universo. Col passare del tempo le disuniformità si contrassero gravitazionalmente formando galassie, e nel loro interno nubi più piccole collassarono gravitazionalmente formando stelle e altri oggetti. Tutti gli oggetti di una galassia sono soggetti anche a un'attrazione gravitazionale netta diretta verso il suo nucleo. Anche se la rotazione della galassia fa rallentare la caduta della materia, ed esplosioni occasionali (per esempio le supernove) possono espellere materia dal centro, le forze di attrito fanno sì che il centro raggiunga inevitabilmente una densità di massa molto più elevata di quella delle regioni esterne della galassia. Gli oggetti estremamente luminosi noti come quasar, che sono osservabili a distanze enormi, possono essere l'estremo risultato di questo processo. Ci si potrebbe attendere, però, che anche il centro di una galassia a spirale di aspetto normale come la nostra possa ospitare quasi tutto lo zoo astronomico galattico: stelle di sequenza principale, giganti rosse, nane bianche, nebulose gassose, nubi di polvere, stelle di neutroni e presumibilmente anche un buco nero - tutti impaccati insieme in uno sciame serrato.

Probabilmente i nuclei galattici sono le sorgenti delle enormi emissioni di materia che si sono osservate in alcune galassie lontane. Sembra proprio che l'energia necessaria per espulsioni così imponenti possa venire solo dai densi nuclei galattici. La nostra galassia mostra anche evidenza di passati eventi esplosivi, anche se relativamente poco violenti, nelle grandi nubi di gas di cui gli astronomi hanno osservato il moto di allontanamento dal nucleo.

Forse le scoperte odierne più entusiasmanti sul nucleo della nostra galassia sono quelle relative al suo parsec centrale, una regione di soli 3,26 anni luce di diametro. Alla nostra distanza dal centro un parsec sottende un angolo di 20 secondi d'arco, che è pari all'angolo sotteso da un cratere lunare di medie dimensioni visto dalla Terra. All'interno di quel volume, che potrebbe stare comodamente nello spazio che separa il Sole dalle stelle più vicine, si trovano milioni di stelle e tutta una varietà di altri oggetti, di cui gli astronomi incominciano solo ora a comprendere la natura. Ci sono anche prove che indicano la presenza di una quantità di materia pari a qualche milione di masse solari concentrata proprio nel cuore del parsec centrale, forse sotto forma di un buco nero. Anche se il buco nero non è osservabile direttamente, la sua esistenza è suggerita dal comportamento del materiale circostante.

Ho scelto l'unità di un parsec per questa trattazione non per qualche suo particolare merito rispetto a qualche altra unità di misura, per esempio l'anno luce, ma semplicemente perché il parsec è l'unità di misura di distanza abituale per gli astronomi. Un parsec è la distanza tra la Terra e un oggetto la cui posizione, osservata dalla Terra, descrive attorno alla sua posizione media un'oscillazione di ampiezza pari a un secondo d'arco (1/3600 di grado) per effetto della parallasse, man mano che la Terra viaggia attorno al Sole. La distanza che ci separa dal centro della galassia è di circa 10 000 parsec, cioè circa 20 volte maggiore della distanza che ci separa dalla famosa Grande Nebulosa in Orione. D'altra parte, il centro della Galassia è circa 70 volte più vicino al sistema solare di quanto sia il nucleo di qualsiasi altra galassia. Pertanto i dettagli del nucleo della nostra galassia dovrebbero risultare visibili 70 volte meglio di quelli del nucleo di qualsiasi altra galassia. Naturalmente c'è il problema che un telescopio puntato verso il centro della nostra galassia non è in grado di distinguere tra gli oggetti in primo piano, quelli nel nucleo o quelli sullo sfondo. Ci sono, tutta-

Il centro della Galassia sta nascosto nella zona meridionale della Via Lattea in direzione della costellazione del Sagittario. La distanza del sistema solare dal centro della Galassia è pari a circa 10 000 parsec, poco più di 30 000 anni luce. La luce dei milioni di stelle che stanno nel nucleo della Galassia è oscurata dalla polvere interstellare. L'emissione radio e infrarossa dal nucleo galattico risulta, però, rivelabile. Il rettangolo, di lunghezza circa doppia del diametro della luna piena, individua l'area che è mostrata nella mappa infrarossa in basso.

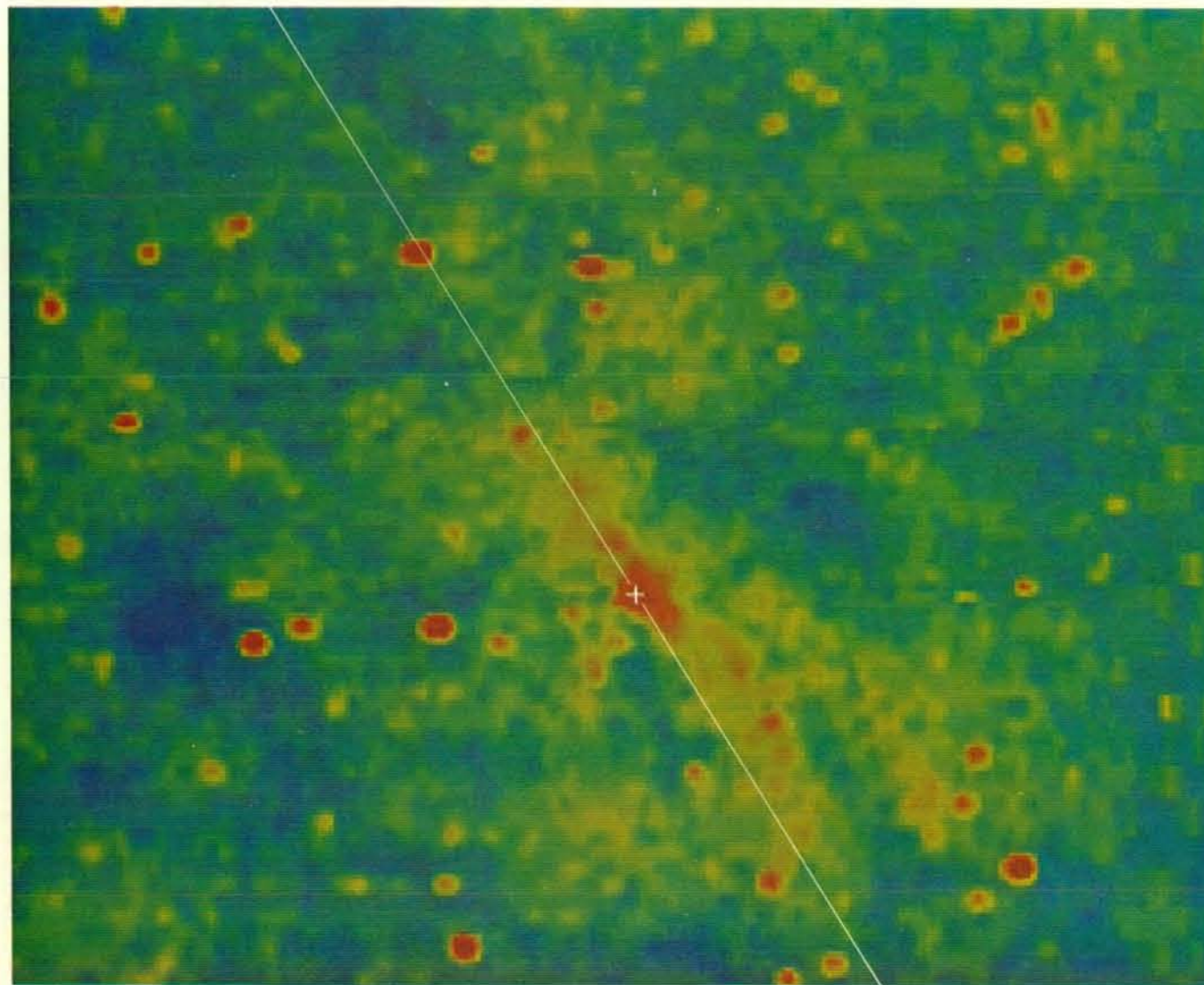



Immagine del centro galattico in infrarosso ottenuta a una lunghezza d'onda di 2,2 micrometri (il quadruplo della lunghezza d'onda della luce verde) da Eric E. Becklin e Gerry Neugebauer del California Institute of Technology e degli Hale Observatories. Il loro strumento di osservazione era il telescopio da un metro di Las Campañas in Cile. In questa immagine a colori codificati, generata dall'elaboratore elettronico, l'intensità della radiazione cresce dal blu al rosso. La crocetta indica il centro della Galassia. La radiazione infrarossa tende a concentrarsi lungo l'equatore galattico, indicato dalla linea bianca sottile. A 2,2 micrometri la principale sorgente della radiazione è costituita da stelle rosse fredde. Le macchie brillanti singole rappresentano giganti rosse in primo piano. L'aspetto irregolare è causato dalla polvere interstellare.

via, vari stratagemmi per operare queste distinzioni e per decidere se una struttura appartiene effettivamente al centro galattico o meno.

I più grandi telescopi e le emulsioni fotografiche più sensibili non possono registrare che una parte estremamente ridotta della luce visibile emessa dai milioni di stelle che si addensano nel nucleo della nostra galassia. A causa delle nubi di polvere interposte solo un fotone di luce visibile su circa 10^{11} fotoni sopravvive nel viaggio di 30 000 anni luce dal centro galattico fino alla Terra. I pochi fotoni superstiti sono poi sommersi dalla radiazione di fondo che limita l'esposizione e che è costituita da fotoni visibili diffusi nell'atmosfera terrestre. Nella regione infrarossa dello spettro, però, la frazione di fotoni che completa il suo lungo viaggio è molto maggiore. Nella banda radio i fotoni provenienti dal nucleo galattico sono deviati e assorbiti solo raramente. All'estremità opposta dello spettro, nella regione dei raggi X, i fotoni viaggiano anch'essi liberamente, ma non sopravvivono al passaggio attraverso l'atmosfera terrestre, così che possono essere registrati solo da strumenti portati a quote elevate. La nostra atmosfera, però, è trasparente per la maggior parte delle lunghezze d'onda radio e per alcune delle infrarosse. A seguito di ciò la maggior parte delle nostre conoscenze sulla regione centrale della Galassia è stata fornita da radiotelescopi e da rivelatori infrarossi estremamente sensibili posti nel fuoco di telescopi a riflessione.

Se la nostra galassia fosse priva di polvere, si potrebbe usare un grande telescopio ottico per registrare tutto il parsec centrale della Galassia su una sola lastra fotografica con una risoluzione di circa un secondo d'arco. Con un telescopio a bordo di un satellite artificiale la risoluzione sarebbe migliorata di un fattore 10 o 20. Questo miglioramento potrebbe essere sufficiente per risolvere circa 100 000 oggetti stellari in un diametro di 20 secondi d'arco, corrispondente al parsec centrale della Galassia. Se però il parsec centrale contiene effettivamente qualche milione di oggetti stellari, come oggi sembra probabile, essi si confonderebbero in addensamenti luminosi, e forse in un'unica massa brillante, anche se fossero fotografati con alta risoluzione da un satellite artificiale.

Ci si può fare un'idea di come potrebbe apparire il nucleo galattico in assenza di polvere guardando fotografie del nucleo della Grande Nebulosa in Andromeda, la galassia a spirale più vicina simile alla nostra. Per via dell'orientamento delle due galassie nello spazio il nucleo della galassia di Andromeda può essere fotografato lungo una linea di vista per lo più libera dalla polvere, che si trova maggiormente concentrata invece nei piani centrali dei due sistemi.

Il fatto che il centro della nostra galassia sia invisibile nell'ottico ha stimolato indubbiamente gli sforzi per osservarlo a lunghezze d'onda diverse. Per ottenere una «immagine» nel radio o nell'infrarosso si devono analizzare strisce successive di cielo (per esempio muovendo il telescopio), registrando punto per punto il numero di fotoni che colpiscono il rivelatore. In questo modo si può costruire un'immagine analoga a quella della televisione con una risoluzione data semplicemente dall'area di cielo che il rivelatore vede in ogni istante. Il limite sulla risoluzione è posto dall'apertura effettiva del telescopio: maggiore il rapporto tra apertura e lunghezza d'onda, più elevata la massima risoluzione raggiungibile. Nella maggior parte dei casi i dati così acquisiti sono trasformati in mappe a curve di livello, nelle quali le linee connettono punti di uguale luminosità nello stesso modo in cui le linee di una mappa topografica connettono punti che si trovano alla stessa quota. Recentemente con l'aiuto dell'elaboratore elettronico è diventato piuttosto facile trasformare questi dati in immagini bidimensionali a colori codificati.

Prima di procedere alla trattazione del parsec centrale della Galassia descriverò brevemente ciò che si è appreso dagli studi nel radio e nell'infrarosso sulla regione nucleare estesa di cui esso fa parte. Sono state ottenute mappe radio a curve di livello della regione a molte lunghezze d'onda. Tutte le mappe rivelano una distribuzione di radiazione allungata attorno all'equatore galattico. La posizione del picco più intenso, noto ai radioastronomi come Sagittario *A*, è pure comune a tutte le mappe e già da tempo si pensava che coincidesse col centro galattico. Sagittario *A* deve avere contribuito al sibilo stazionario osservato da Jansky, anche se il suo strumento mancava della risoluzione necessaria per distinguerlo da altri picchi vicini.

Mappa radio del centro galattico ottenuta a una lunghezza d'onda di 3,75 centimetri da Dennis Downes e Arthur E. Maxwell della Harvard University e da M. L. Meeks del Massachusetts Institute of Technology col radiotelescopio da 36,6 metri di Westford, Massachusetts. Il gas ionizzato è la principale sorgente di emissione radio. Il rettangolo racchiude la regione nell'immagine in infrarosso della pagina precedente. Il gas ionizzato e le stelle sono concentrate lungo l'equatore galattico, e hanno la massima luminosità in corrispondenza della radiosorgente Sagittario *A*.

Le sorgenti di radiazione nel nucleo galattico comprendono stelle, polvere riscaldata e le nubi di gas ionizzato note come regioni H II. Il termine H II sta a indicare atomi di idrogeno che sono stati dissociati in protoni ed elettroni dalla radiazione ultravioletta emessa da stelle vicine di alta temperatura. Gli elettroni e i protoni emettono radiazione su un ampio intervallo di lunghezza d'onda (la radiazione continua) quando sono accelerate da forze di natura elettromagnetica, prime tra tutte sono da annoverare quelle che esercitano gli uni sugli altri. La curva indica la quantità di radiazione continua che sopravvive al passaggio attraverso la polvere interstellare per giungere agli strumenti di rilevazione posti sulla Terra. Prima dello sviluppo dell'astronomia radio e infrarossa il nucleo galattico era assolutamente inaccessibile alle osservazioni.

Le sorgenti di emissione che appaiono sulle mappe radio del centro galattico sono nubi di gas caldo interstellare note come regioni H II. Queste regioni, comuni nella nostra galassia, sono costituite da atomi ionizzati, cioè da atomi che sono stati privati di uno o più elettroni. Gli atomi, per lo più di idrogeno, sono ionizzati dalla radiazione ultravioletta proveniente da stelle vicine (o da altre sorgenti), che sono molto più calde del Sole. I frammenti atomici carichi positivamente e gli elettroni liberi carichi negativamente di una regione H II emettono onde radio in un ampio intervallo di lunghezze d'onda quando sono accelerati, per esempio in quasi-collisioni con altre particelle cariche. Al contrario, il gas elettricamente neutro, come quello delle regioni H I (che consiste soprattutto di idrogeno in forma atomica) o delle nubi molecolari (costituite soprattutto da idrogeno in forma molecolare), emette poca energia nel radio tranne che a limitate lunghezze d'onda discrete.

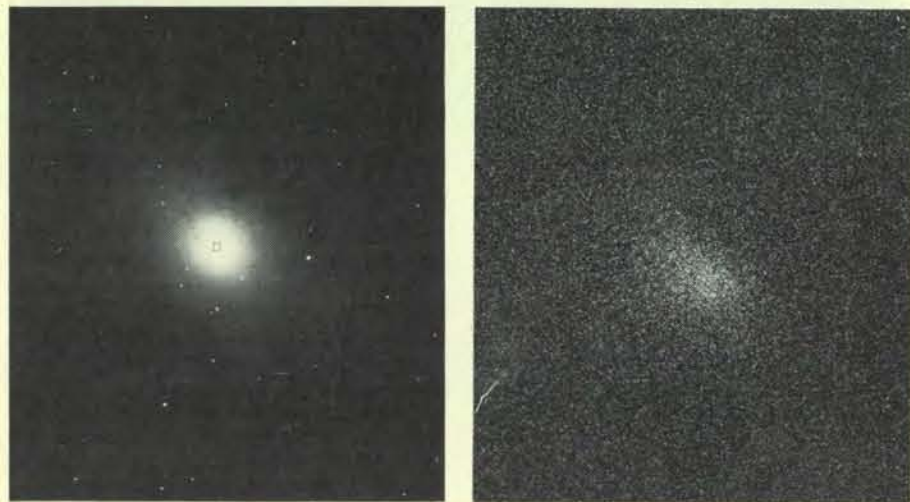
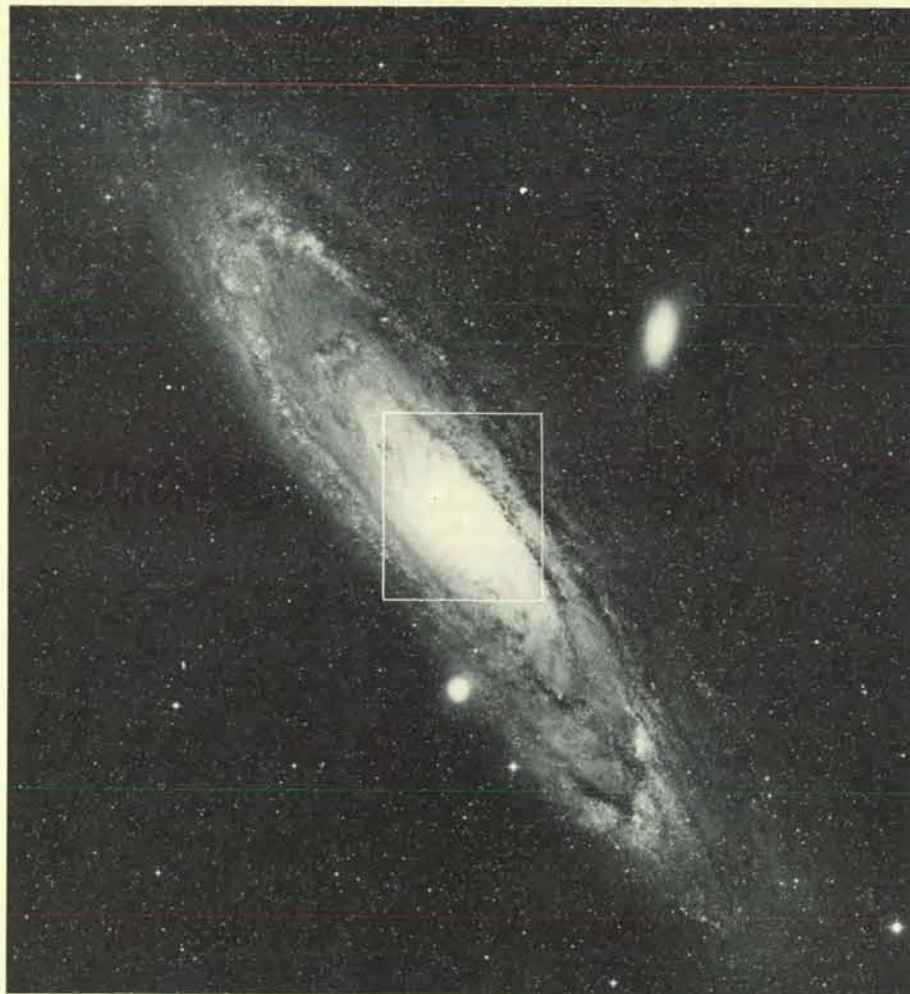
Un quadro diverso della regione nucleare estesa emerge da osservazioni fatte nell'infrarosso a una lunghezza d'onda di 2,2 micrometri, pari circa al quadruplo della lunghezza d'onda della radiazione visibile. A differenza dell'emissione radio che proviene dal gas interstellare ad alta temperatura, la radiazione a 2,2 micrometri è costituita essenzialmente da luce stellare. La maggior parte della radiazione a 2,2 micrometri è emessa da giganti rosse, che sono stelle luminose di grandi dimensioni in uno stadio evolutivo avanzato con temperature superficiali di soli 2000 kelvin, basse rispetto ai 6000 kelvin del Sole. Anche se le giganti rosse non possono rappresentare che una piccola percentuale delle stelle del centro galattico, dovrebbero essere distribuite in modo abbastanza uniforme tra le stelle più giovani e brillanti, e perciò dovrebbero fornire in linea di principio una buona indicazione della distribuzione di tutte le stelle circostanti.

Un'immagine a colori codificati della regione centrale estesa della Galassia a 2,2 micrometri è riportata nella illustrazione in basso di pagina 43. Parte dell'aspetto irregolare della immagine si spiega con le variazioni nella quantità di polvere interstellare che si frappone tra il sistema solare e le diverse parti del centro galattico. Gli oggetti puntiformi più luminosi nell'immagine sono probabilmente giganti rosse e sono certamente oggetti in primo piano. Questa deduzione è sostenuta dall'uniformità della distribuzione sull'immagine.

Pure con le eccezioni già dette è chiaro che l'emissione a 2,2 micrometri, come l'emissione radio, è concentrata lungo l'equatore galattico. La direzione del massimo di luminosità a 2,2 micrometri coincide con quella di un picco radio all'interno di Sagittario *A*, noto come Sagittario *A* Ovest. Anche le mappe infrarosse a lunghezze d'onda più elevate, come quelle emesse dalla stessa polvere interstellare, presentano il loro picco più brillante in prossimità di Sagittario *A* Ovest. Dunque, le concentrazioni di stelle, di gas ad alta temperatura e di polvere a temperatura più bassa raggiungono il loro massimo tutte nella stessa direzione. Possono restare pochi dubbi sul fatto che lungo quella direzione si trova il centro vero e proprio della Galassia.

Dalle osservazioni a 2,2 micrometri, che furono fatte da Eric E. Becklin e Gerry Neugebauer del California Institute of Technology e degli Hale Observatories, si può stimare la massa delle stelle nel centro della Galassia. Becklin e Neugebauer trovarono che l'intensità e la distribuzione dell'emissione a 2,2 micrometri dal centro della galassia di Andromeda, indicata come M 31, sono paragonabili a quelle dell'emissione dal centro della nostra galassia. Dato che il nucleo di M 31 può essere fotografato chiaramente nel visibile, si può stimare con ragionevole precisione il numero di stelle che devono essere presenti. Robert H. Sanders e Thomas Lowinger della Columbia University e J. H. Oort dell'Università di Leida hanno calcolato la distribuzione della massa stellare nel centro della nostra galassia, nell'ipotesi che il rapporto tra la radiazione infrarossa e la radiazione visibile sia lo stesso che si osserva nel centro di M 31. Essi stimarono che una sfera con un parsec di diametro centrata nella posizione del picco centrale brillante dell'emissione a 2,2 micrometri debba contenere una massa sotto forma di stelle pari a circa due milioni di masse solari.

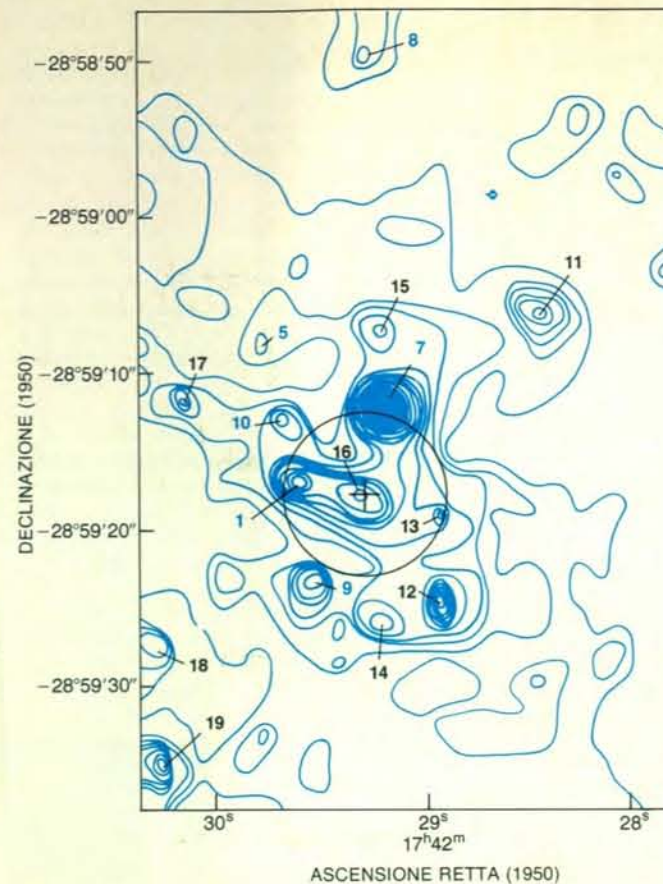
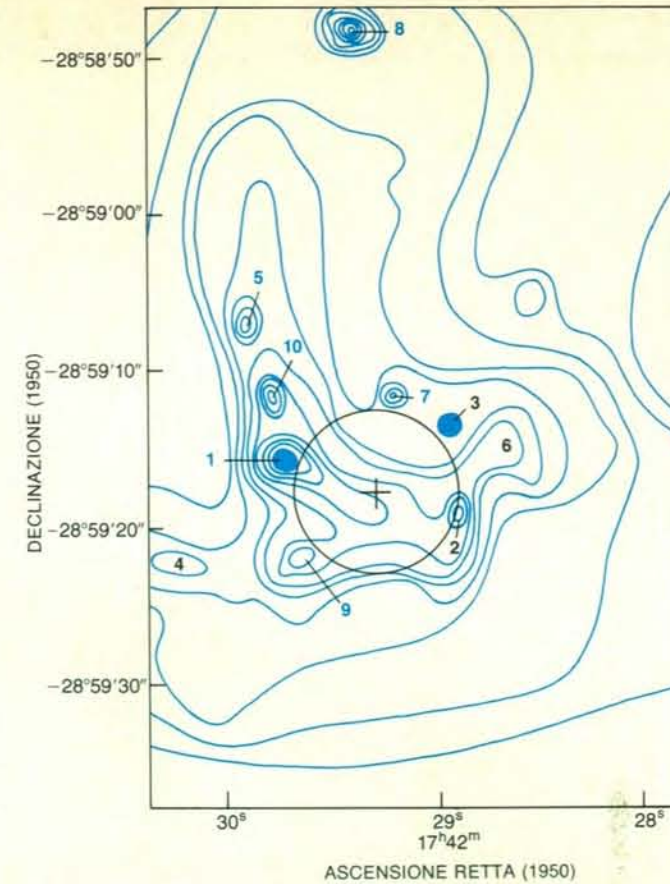
L'ammasso globulare di stelle più densamente impaccato fuori dal nucleo galattico contiene circa 10 000 stelle nel suo parsec centrale. Entro un parsec di distanza dal Sole non c'è nessun'altra stella. Se si suppone che tutte le stelle nel parsec centrale siano simili al nostro Sole, la stima precedente ci dice che la distanza media tra le stelle nel parsec centrale è di circa una settimana luce, cioè circa 100 volte la distanza che separa il Sole dal pianeta Plutone. Se la Terra si trovasse nel parsec centrale, i milioni di stelle circostanti la inonderebbero con una luce pari a quella di varie centinaia di lune piene. Il fatto che la vita possa o meno esistere nel parsec centrale pieno di radiazione è un altro problema.

Osservando una regione piccola e lontana come il parsec centrale della Galassia sarebbe desiderabile lavorare con strumenti dotati della massima risoluzione angolare possibile. I più grandi radiotelescopi operanti a piccole lunghezze d'onda raggiungono al più una risoluzione di circa un minuto d'arco, una risoluzione così scarsa da confondere qualsiasi struttura spaziale del parsec centrale. La risoluzione, però, può essere aumentata

Il nucleo di una galassia a spirale vicina, la Grande Nebulosa in Andromeda, nota come Messier 31 (M31), può essere fotografato nell'ottico, benché sia 70 volte più distante del nucleo della nostra galassia, perché è poca la polvere interposta che impedisce il passaggio della luce. In M31, come nel nostro sistema a spirale, la polvere è concentrata lungo il piano del disco galattico. Le due galassie sono orientate nello spazio in modo tale che possiamo vedere il nucleo di M31 senza ostacoli. Nella fotografia in alto, che mostra l'intera galassia, la regione centrale è fortemente sovraesposta. La regione contenuta nel rettangolo è rappresentata nella fotografia in basso a sinistra: un'esposizione di un minuto col telescopio Hale da cinque metri di Monte Palomar. Anche se la maggior parte delle stelle che riempiono il rettangolo della fotografia in alto è ora invisibile, le miriadi di stelle strettamente impaccate nei 100 parsec centrali sono ancora sovraesposte. La fotografia in basso a destra mostra i 20 parsec centrali del nucleo, l'area racchiusa nel rettangolo sulla seconda fotografia. La regione irrisolta nella terza fotografia, circa cinque parsec di estensione, rappresenta la luce di milioni di stelle; i puntini bianchi esterni sono grani dell'emulsione fotografica molto ingranditi. La fotografia è stata fatta con un telescopio da 20 centimetri posto a bordo di un pallone stratosferico da E. S. Light, R. E. Danielson e Martin Schwarzschild dell'Università di Princeton. Le osservazioni infrarosse suggeriscono che le stelle del nucleo di M31 e quelle del nucleo della nostra galassia sono distribuite in modo simile.

notevolmente usando due o più radiotelescopi separati da una grande distanza e operando con tecniche interferometriche. Sfortunatamente la posizione meridionale del centro galattico lo ha reso quasi inaccessibile all'interferometria coi grandi radiotelescopi, dato che fino a poco tempo fa ne esistevano solo a latitudini settentrionali elevate. La situazione migliorerà presto, però, col completamento previsto per l'anno prossimo della grande schiera radio in allestimento presso Socorro, nel Nuovo Messico. Questo strumento consentirà finalmente di esplorare il parsec centrale nel radio con una risoluzione angolare paragonabile a quella che si ottiene oggi nell'infrarosso.

L'elevata risoluzione resa possibile a lunghezze d'onda radio dall'interferometria è già stata applicata, seppure in modo limitato, allo studio di Sagittario *A* Ovest. Nel 1974 Bruce Balick, allora del Lick Observatory e Robert L. Brown del National Radio Astronomy Observatory di Green Bank, nella Virginia occidentale, scoprirono per mezzo dell'interferometria radio che Sagittario *A* Ovest contiene in prossimità del suo centro un oggetto radioemittente che è nello stesso tempo estremamente piccolo ed estremamente luminoso. In seguito si è dimostrato che questa radiosorgente compatta ha un nucleo centrale con un diametro non più grande di 0,001 secondi d'arco, pari solo a 10 volte la distanza Terra-Sole (nell'ipotesi che l'oggetto si trovi effettivamente nel centro della Galassia).

Nonostante che l'oggetto radio compatto presenti qualche somiglianza con certe specie di stelle radioemittenti, soddisfa anche alcuni requisiti sul modo in cui si rivelerebbe la materia quando è attirata in un buco nero. Si pensa che un buco nero possa originarsi quando una stella massiccia, che ha esaurito il suo combustibile nucleare, non può più generare la pressione di radiazione necessaria per controbilanciare l'attrazione verso l'interno della gravità. In queste condizioni l'interno stellare collasserà fino a diventare così piccolo e denso che le leggi fisiche conosciute non possono più descriverlo. La forza di attrazione gravitazionale in prossimità di un simile oggetto collassato è così grande che attorno a esso c'è una regione da cui non può sfuggire assolutamente nulla, nemmeno la luce; questa regione costituisce un buco nero.

La materia che viene a trovarsi nell'intenso campo gravitazionale appena fuori da un buco nero sarà accelerata fino a raggiungere velocità elevate e sarà riscaldata dalle collisioni fino ad altissime temperature. A seguito di ciò un buco nero dovrebbe essere circondato da una piccola ma intensa sorgente di emissione radio di forma discoidale, così da assomigliare a una radiosorgente compatta. Nel centro galattico, dove c'è abbondanza di gas e polvere, un buco nero dovrebbe inghiottire materia con un tasso relativamente elevato, accrescendo rapidamente la propria massa e la propria influenza gravitazionale. La possibilità che Sagittario *A* Ovest ospiti un buco nero massiccio nel

Mappe infrarosse del centro galattico ottenute a due lunghezze d'onda differenti, 2,2 micrometri *(a sinistra)* e 10,6 micrometri *(a destra)*, presentano strutture diverse. Il cerchio sulle due mappe rappresenta quello che si pensa sia il parsec centrale della Galassia. Solo sei delle sorgenti infrarosse più brillanti appaiono su entrambe le mappe: 1, 5, 7, 8, 9 e 10. Alcune sorgenti luminose a 2,2 micrometri sono quasi certamente giganti rosse, stelle fredde e molto luminose. IRS (che sta per «sorgente infrarossa») 7, IRS 11 e IRS 12 sono giganti rosse. Si pensa che la maggior parte delle sorgenti sulla mappa a 10,6 micrometri, IRS 3 e IRS 7 escluse, siano nubi di polvere calde associate a regioni H II. La croce sulle mappe coincide con la radiosorgente brillante all'interno di Sagittario *A* Ovest, che circonda forse un buco nero. La mappa a 2,2 micrometri è stata ottenuta da Becklin e Neugebauer col telescopio da cinque metri. La mappa a 10,6 micrometri è un mosaico di mappe ottenute con telescopi diversi da G. H. Rieke dell'Università dell'Arizona e da C. M. Telesco e D. A. Harper dell'Università di Chicago.

centro della Galassia è entrata senza dubbio nei pensieri di molti astronomi. Nel seguito esamineremo prove recenti di questa possibilità.

L'esame dettagliato dei parsec centrali della Galassia è stato molto più facile nell'infrarosso che nel radio. Negli ultimi sei anni sono stati fatti studi in infrarosso a varie lunghezze d'onda, soprattutto da G. H. Rieke e Frank J. Low dell'Università dell'Arizona e da Becklin e Neugebauer. A tutte le lunghezze d'onda hanno trovato una concentrazione brillante di oggetti entro un parsec di diametro coincidente con Sagittario *A* Ovest. La luminosità relativa e la distribuzione degli oggetti risultano, però, decisamente diverse a lunghezze d'onda differenti.

Le osservazioni a 2,2 micrometri e a 10 micrometri mostrano forse al meglio queste differenze *(si vedano le illustrazioni di questa pagina)*. Sia all'una che all'altra lunghezza d'onda si vedono molti oggetti distinti, ma solo sei di loro appaiono a entrambe le lunghezze d'onda. In altre parole, oggetti che spiccano a una lunghezza d'onda, non fanno altrettanto nell'altra. Per esempio, l'oggetto noto come IRS 1 (*infrared source 1* = sorgente infrarossa 1), che è il corpo più brillante a 10 micrometri, è solo moderatamente luminoso a 2,2 micrometri. IRS 7, che è l'oggetto di gran lunga più brillante a 2,2 micrometri, è appena visibile a 10 micrometri. Sembrerebbe perciò che gli oggetti che si osservano a lunghezze d'onda diverse, appartengano a classi differenti.

Che specie di oggetti sono? Tutti brillano nell'infrarosso con un'intensità superiore a quella che avrebbe la grande maggioranza delle stelle a una distanza di 10 000 parsec. Dato che, come ho ricordato, le giganti rosse sono particolarmente luminose a 2,2 micrometri, è naturale chiedersi se qualche sorgente brillante a 2,2 micrometri sia una gigante rossa. C'è un modo per rispondere a questa domanda. Nelle atmosfere delle stelle più fredde si possono formare molecole. Una specie che si forma in abbondanza è quella dell'ossido di carbonio, che assorbe radiazione a lunghezze d'onda maggiori o uguali a 2,3 micrometri. L'assorbimento è particolarmente intenso nelle giganti rosse brillanti. Nel 1975 osservazioni condotte da gruppi dell'Università dell'Arizona e del Caltech (California Institute of Technology) rivelarono strutture di assorbimento a 2,3 micrometri nello spettro di tre sorgenti in prossimità del centro galattico (IRS 7, IRS 11 e IRS 12), indicando così che ciascuna delle tre sorgenti contiene almeno una gigante rossa luminosa. D'altra parte, una delle sorgenti brillanti a 2,2 micrometri, IRS 16, presenta poco o niente dell'assorbimento dell'ossido di carbonio. L'identificazione di IRS 16 resta un problema aperto. La sua posizione sembra coincidere con quella della radiosorgente compatta in Sagittario *A* Ovest, e pertanto è una sorgente particolarmente interessante.

Un altro oggetto che non presenta assorbimento da ossido di carbonio, IRS 1, fa parte della classe di oggetti che sono molto più luminosi a 10 micrometri che a 2,2. Ben poco dell'emissione a 10 micrometri di oggetti di questa classe può essere radiazione stellare diretta, perché anche le stelle più fredde sono molto più deboli a 10 micrometri che a 2,2. Gli astronomi che operano nell'infrarosso hanno trovato ora molte sorgenti distribuite qua e là per la Galassia che risultano insolitamente brillanti a 10 micrometri. In quasi tutti i casi si è identificata nella polvere calda la sorgente della radiazione. Le particelle di polvere assorbono la

radiazione visibile e ultravioletta emessa da stelle vicine o da altri oggetti luminosi e riirradiano nell'infrarosso l'energia assorbita. Probabilmente nel nucleo galattico accade qualcosa di simile, e si può assumere che i picchi sulla mappa a 10 micrometri rappresentino nubi di polvere che si trovano nel parsec centrale o nelle sue vicinanze.

Il gas interstellare nella nostra galassia è circa 100 volte più abbondante in massa della polvere interstellare; inoltre, la presenza di polvere implica sempre la presenza del gas. Pertanto, sembra ragionevole pensare che il gas radioemittente di Sagittario *A* Ovest sia gas associato a nubi di polvere, e che sia riscaldato dalla stessa sorgente (o sorgenti) che riscalda queste nubi. Se le cose stanno così, ci si aspetterebbe che mappe radio ad alta risoluzione di Sagittario *A* Ovest rivelino una struttura simile a quella osservata nelle mappe a 10 micrometri. Nel 1974 ci furono indicazioni da studi radiointerferometrici a favore del fatto che Sagittario *A* Ovest fosse costituito da un insieme di nubi di gas ionizzato più compatte. A quel tempo, però, era difficile determinare se ci fossero o no associazioni spaziali tra i picchi di emissione delle mappe radio e di quelle infrarosse.

La trattazione sviluppata fin qui copre approssimativamente lo stato delle nostre conoscenze sul nucleo galattico nel 1975. Si sapeva che nel parsec centrale c'era una concentrazione massiccia di stelle, la cui luce visibile è nascosta dalla polvere. Entro qualche parsec dal centro erano state trovate sorgenti infrarosse di vari tipi. Si era dimostrato che alcune sorgenti brillanti a 2,2 micrometri erano giganti rosse fredde e luminose. La maggior parte delle altre sorgenti potevano ospitare anche esse stelle fredde, ma queste non erano state ancora identificate. La maggior parte degli oggetti luminosi a 10 micrometri assomigliava a regioni H II compatte, ma, anche se si sapeva che il gas ionizzato era presente nel centro galattico, non si sapeva ancora se il gas era associato fisicamente a qualcuna di queste sorgenti infrarosse. Si era trovato che una sorgente interessante di emissione radio, probabilmente un oggetto supermassiccio proprio nel centro della Galassia, coincideva con una sorgente brillante peculiare che emette intensamente intorno a 2,2 micrometri. Infine, non si sapeva ancora assolutamente nulla sui moti di tutti questi oggetti; la materia nel parsec centrale avrebbe potuto collassare, espandersi, ruotare oppure muoversi casualmente.

Le stelle, il gas e la polvere nel centro galattico emettono radiazione la cui intensità varia in modo piuttosto regolare al variare della lunghezza d'onda; pertanto questa radiazione è detta continua. Un'altra sorgente di informazione potenzialmente ricca è l'emissione di righe: cioè, la radiazione emessa e assorbita da atomi e molecole a frequenze specifiche e caratteristiche. Di solito lo spettro di un oggetto astronomico è costituito sia da un continuo regolare che da righe. Ogni riga è associata a uno specifico stato energetico di un atomo o di una molecola ed è caratterizzata da un'intensità, una lunghezza d'onda centrale e una larghezza.

L'intensità di una riga è una misura della temperatura e dell'abbondanza della sostanza emittente. Se la sorgente si muove verso l'osservatorio, la riga è spostata a una lunghezza d'onda minore; se la sorgente si sta allontanando, la riga è spostata a una lunghezza d'onda maggiore. Questi spostamenti Doppler danno indicazioni all'astronomo sulla velocità radiale: la velocità di una sorgente lungo la linea di vista. Velocità dirette verso l'osservatore («spostamenti verso il blu») sono dette negative; quelle di allontanamento dall'osservatore («spostamenti verso il rosso») sono dette positive. Quando gli atomi o le molecole della sorgente si muovono a velocità diverse, gli spostamenti variano. La riga risulta allargata e pertanto si dice che ha una larghezza. Dunque, la larghezza di una riga può indicare l'intervallo di velocità radiali presenti nella sorgente.

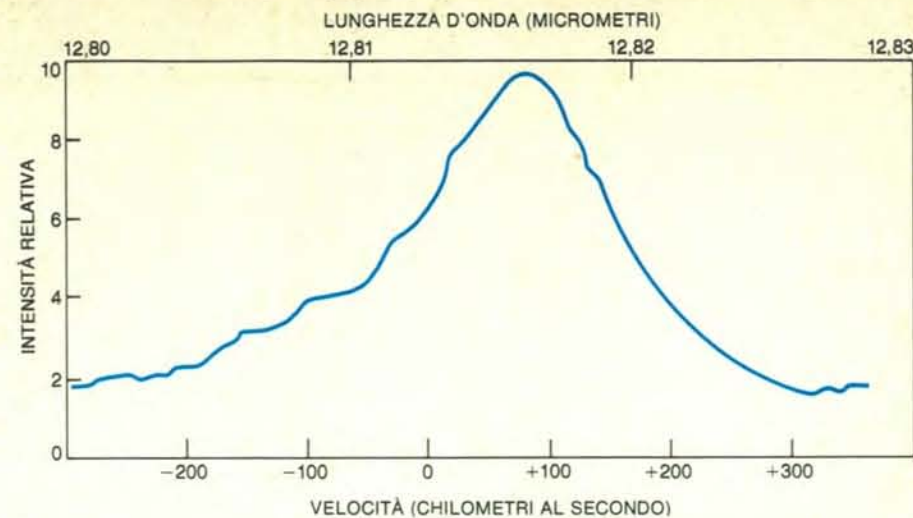
Nel 1973 David Aitken, Barbara Jones e James Penman dell'University College di Londra riuscirono a rivelare per la prima volta una riga di emissione nell'infrarosso generata nel centro galattico. La lunghezza d'onda della riga, circa 12,8 micrometri, consentì di identificarne la sorgente nel neo ionizzato una sola volta, cioè Ne II: neo mancante di uno dei suoi 10 elettroni. In precedenza si era già trovato che quella stessa riga del neo era una delle righe di emissione infrarossa più brillanti in varie regioni ionizzate, altrove nella Galassia. Queste regioni si rendono manifeste anche per mezzo della radiazione continua che emettono nella regione radio dello spettro.

Aitken, Jones e Penman hanno calcolato l'intensità prevista per la riga del neo facendo uso del valore conosciuto per l'intensità dell'emissione radio da Sagittario *A* Ovest e supponendo che l'abbondanza del neo nel gas ionizzato di Sagittario *A* Ovest sia normale. L'intensità osservata della riga era leggermente superiore al valore calcolato da loro, indicando così che il gas nel centro galattico è più ricco di neo del gas osservato altrove. Il valore calcolato era però abbastanza vicino a quello osservato da indicare che il medesimo gas ionizzato produce sia la riga infrarossa del neo che l'emissione continua nel radio. Il lavoro di Londra ha così dimostrato che la riga a 12,8 micrometri del Ne II, che costituisce solo una frazione ridotta (approssimativamente lo 0,01 per cento) del gas ionizzato, poteva essere usata come indicatore nello studio del moto e della distribuzione del gas ionizzato di Sagittario *A* Ovest, che si pensava si trovasse proprio nel centro della Galassia.

Spettri infrarossi di cinque sorgenti in prossimità del centro galattico. Tre, IRS 7, IRS 11 e IRS 12 presentano il carattere distintivo delle giganti rosse: una caduta di intensità a lunghezze d'onda maggiori di 2,3 micrometri. Esse hanno una atmosfera ricca di ossido di carbonio, che assorbe la radiazione a 2,3 micrometri e oltre. IRS 1, una regione H II, e IRS 16, in Sagittario *A* Ovest, non presentano questo assorbimento. Gli spettri sono stati ottenuti col telescopio da cinque metri da Becklin, Neugebauer, Steven V. W. Beckwith, Keith Matthews e C. G. Wynn-Williams.

Il gruppo di Londra lavorava con uno strumento infrarosso con un campo circolare di diametro pari approssimativamente a un parsec alla distanza del centro galattico. La risoluzione del loro spettrometro era troppo bassa per determinare la larghezza della riga del neo o per stabilire se essa mostrava spostamenti Doppler. Nel dipartimento di fisica dell'Università della California a Berkeley, ove mi trovavo come studente di dottorato, una serie di studenti che lavorava con Charles H. Townes aveva sviluppato spettrometri infrarossi dotati di elevata risoluzione spettrale. Nel 1975 Eric Wollman, John Lacy e io, insieme con Townes e David Rank dell'Università della California a Santa Cruz, misurammo la forma della riga del Ne II utilizzando il telescopio da 1,5 metri dell'Osservatorio Inter-Americano di Cerro Tololo in Cile (dove la costellazione del Sagittario passa proprio sulla verticale). Ciò che trovammo mostrò chiaramente che Sagittario *A* Ovest non è una normale regione H II galattica. La riga del neo aveva una larghezza di varie centinaia di chilometri al secondo, mentre le velocità per una regione H II sono comprese normalmente nell'ambito di un intervallo di 30 chilometri al secondo.

Quasi contemporaneamente osservazioni radio di righe di emissione generate dalla ricombinazione di nuclei di idrogeno con elettroni operate da Thomas Pauls, Peter G. Mezger ed Edward Churchwell del Max Planck Institut für Radioastronomie di Bonn mostrarono pure che il gas ionizzato in Sagittario *A* Ovest è in rapido movimento. Ancora non si sapeva se il moto era ordinato o turbolento e non si conosceva la distribuzione del neo e dell'altro gas ionizzato all'interno del parsec centrale.

Le risposte a queste e ad altre domande richiedevano osservazioni della riga a 12,8 micrometri con una maggiore risoluzione angolare. Al Lick Observatory nel 1975 e nel 1976 osservammo molte posizioni all'interno del parsec centrale e trovammo che sia l'intensità della riga che la velocità radiale variavano con la posizione. Nel 1977 Lacy completò un nuovo spettrometro più sensibile che consentiva di osservare la riga del Ne II ancora più dettagliatamente. Nel 1977 e nel 1978 Lacy, Townes, Fred Baas (di Leida, ma temporaneamente con noi) e io eseguimmo una misurazione completa della riga del Ne II sul parsec centrale e oltre lavorando con questo strumento al nuovo telescopio Du Pont da 2,5 metri dell'Osservatorio di Las Campañas in Cile. Operando con diametri di apertura corrispondenti a solo 0,2 parsec alla distanza del centro galattico, abbiamo ottenuto spettri in un reticolo rettangolare di posizioni e abbiamo costruito mappe a curve di livello dell'intensità della riga a velocità differenti.

L'emissione dal neo ionizzato una sola volta, Ne II, presente nelle nubi di gas nel parsec centrale della Galassia, fu studiata dall'autore e dai suoi colleghi con uno spettrometro infrarosso connesso al telescopio da 1,5 metri dell'Osservatorio Inter-Americano di Cerro Tololo in Cile. L'esame della riga di emissione del Ne II a circa 12,8 micrometri rivela che la linea è stata allargata notevolmente dal moto delle nubi di gas, di cui il neo è un costituente minore. Il moto verso l'osservatore (detto negativo) sposta la riga a frequenze più elevate, cioè a minori lunghezze d'onda, mentre il moto di allontanamento dall'osservatore (detto positivo) ha l'effetto opposto. Le velocità osservate del neo ionizzato variano approssimativamente tra —200 e +200 chilometri al secondo lungo la linea di vista. Questo indica la presenza di intense forze gravitazionali presso il centro della Galassia. La intensità della radiazione continua emessa a circa 12,8 micrometri dalla polvere calda del nucleo è data dal livello di emissione a entrambi i lati della riga allargata del neo.

I nostri dati mostrano una grande varietà di situazioni sia nell'intensità della riga del neo che nella velocità del neo osservato in direzione del parsec centrale. In qualche direzione il profilo di intensità della riga presenta picchi a più di una velocità. Chiaramente Sagittario *A* Ovest non può essere considerato una regione uniforme di gas ionizzato, ma deve essere costituito da un certo numero di sorgenti indipendenti più piccole in moto l'una rispetto all'altra. Come ho indicato, gran parte della radiazione continua infrarossa a 2,2 micrometri proviene anch'essa da gruppi di piccole sorgenti. Era importante stabilire se la distribuzione del neo ionizzato è in correlazione con le mappe a 2,2 o a 10 micrometri.

Per studiare questo problema prendemmo nota delle posizioni delle intensità massime della riga del Ne II a varie velocità. Rappresentando le posizioni dei componenti di velocità sulle mappe del continuo infrarosso, trovammo che i massimi di intensità coincidevano con molte sorgenti che erano brillanti a 10 micrometri. Per esempio, IRS 1 si trova nella stessa posizione di un picco della riga del neo a velocità zero. In prossimità di IRS 2 c'è poco neo ionizzato a velocità zero, ma c'è un picco molto netto di neo ionizzato in moto verso di noi a circa 280 chilometri al secondo, una velocità che non si riscontra altrove nel parsec centrale. Almeno altre cinque sorgenti infrarosse, IRS 4, IRS 5, IRS 6, IRS 9 e IRS 10, coincidono con componenti di velocità localizzate della riga del Ne II.

Le coincidenze spaziali di gas ionizzato molto caldo e di polvere calda sono prove convincenti del fatto che la maggior parte delle sorgenti brillanti sulla mappa del continuo a 10 micrometri è costituita da regioni H II compatte all'interno di Sagittario *A* Ovest. Dall'intensità della riga del neo si può stimare che il materiale ionizzato in ciascuna di queste regioni ammonti approssimativamente a una massa solare. La massa della polvere associata, che è responsabile dell'emissione continua a 10 micrometri, è molto minore di quel valore. Normalmente una regione H II compatta ha un diametro di un quarto di parsec. Dato che nel parsec centrale ci sono circa due milioni di stelle, è probabile che ogni regione H II compatta ospiti circa 100 000 stelle.

Solo due sorgenti sulla mappa a 10 micrometri certamente non sono regioni H II. Già si sa che una di esse, IRS 7, è una gigante rossa. L'altra, IRS 3, è più misteriosa. IRS 3 non può essere una gigante rossa, perché non è luminosa alle lunghezze d'onda infrarosse più corte. Non può essere nemmeno una regione H II, perché l'intensità della riga del Ne II diminuisce chiaramente in prossimità della sorgente. Questo fa pensare che IRS 3 oscuri parte del gas ionizzato. Forse IRS 3 è una nube di gas e polvere fredda e opaca davanti al parsec centrale espulsa recentemente da una stella, o che sta collassando gravitazionalmente per diventare una stella.

Tutte le regioni H II compatte nel parsec centrale sono in movimento, e la componente radiale della loro velocità è data dallo spostamento Doppler della riga del neo associata. Per lo più le regioni H II differiscono chiaramente per posizione o per velocità, perciò devono essere

oggetti indipendenti. Ciononostante, lo studio di tutti gli spettri del neo indica la presenza di un moto di insieme nel parsec centrale. Si può notare questo più facilmente osservando che una linea approssimativamente in direzione nord-sud attraverso il centro della regione divide il neo ionizzato in cui predominano moti di avvicinamento dal neo ionizzato in cui predominano moti di allontanamento. L'interpretazione più semplice di questa suddivisione è la rotazione; la linea di divisione corrisponderebbe all'asse di rotazione.

Finora si è trovata evidenza di rotazione solo nel gas ionizzato, ma sembra ragionevole che anche il resto della materia del parsec centrale (stelle comprese) si muova in modo sostanzialmente analogo. Dagli spostamenti Doppler della riga del neo osservati a qualche decina di parsec dall'asse si può stimare in circa 10 000 anni il tempo necessario perché il gas, la polvere e le stelle nel parsec centrale completino un'orbita attorno al centro galattico. (Il Sole impiega circa 200 milioni di anni.) L'orientazione dell'asse è molto particolare: è quasi perpendicolare all'asse di rotazione del disco galattico. Questo fa pensare che la regione del nucleo costituisca una struttura a sé stante, nettamente distinta dall'ambiente circostante. Non si sa ancora fino a quale distanza dal centro si estenda questo asse di rotazione peculiare.

I moti del gas ionizzato nel parsec centrale, indipendentemente dal fatto che siano parte di una rotazione collettiva o meno, sono influenzati dalla forza gravitazionale prodotta dalla massa contenuta nello stesso volume e possono essere usati per calcolare l'ammontare di questa massa. Il resto della Galassia non esercita praticamente nessuna forza gravitazionale netta sul parsec centrale, dato che è distribuito più o meno simmetricamente attorno al centro, con la maggior parte della sua massa a grande distanza da esso. La massa totale del parsec centrale, stimata in base alle velocità della riga del neo, è compresa tra cinque e otto milioni di masse solari. L'analisi delle osservazioni a 2,2 micrometri indica che solo due milioni di masse solari in questo totale sono sotto forma di stelle di qualsiasi tipo. La massa totale della polvere e del gas all'interno di questa regione è trascurabile al confronto: probabilmente non supera 10 masse solari circa.

Pertanto, risulta importante decidere se la differenza, compresa tra tre e sei milioni di masse solari, che si ottiene sottraendo la massa stellare osservata dalla massa totale, è significativa. Rappresenta massa mancante ancora da spiegare? Oppure le stime non sono precise? Un'osservazione può essere pertinente. Misure della riga infrarossa del neo e delle righe radio generate dalla ricombinazione dei nuclei di idrogeno con gli elettroni indicano che le velocità del gas entro pochi parsec dal centro tendono ad aumentare avvicinandosi al centro. È stato fatto notare da Luis Rodriguez e da Eric J. Chaisson del Centro per l'astrofisica del Harvard College Observatory e dello Smithsonian Astrophysical Observatory, e da altri ancora, che questo incremento non poteva essere spiegato solamente con la forza gravitazionale dell'ammasso stellare centrale, a meno che questo ammasso avesse un nucleo insolitamente denso o contenesse probabilmente un oggetto centrale compatto e massiccio presumibilmente non stellare.

Considerando modelli teorici che cercano di spiegare le osservazioni del parsec centrale supponiamo (almeno inizialmente) che il nucleo non sia osservato in un'epoca particolare. Le osservazioni hanno mostrato che il parsec centrale contiene un ammasso estremamente denso di stelle insieme con una ragguardevole quantità di polvere e di gas ionizzato, distribuita in modo non uniforme e in rapido movimento. Un modello soddisfacente dovrebbe essere in grado di spiegare non solo la presenza di queste caratteristiche, ma anche la loro persistenza.

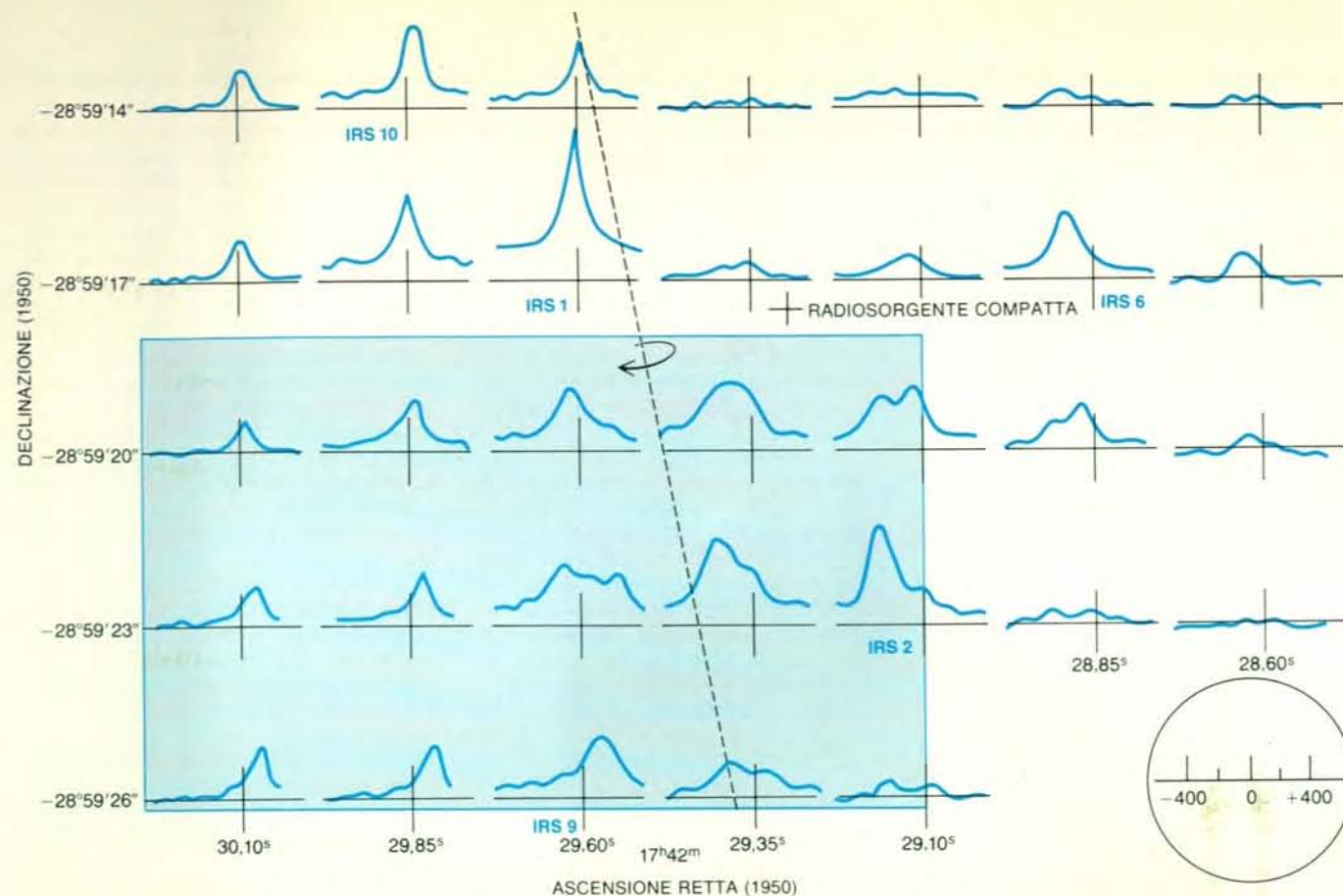
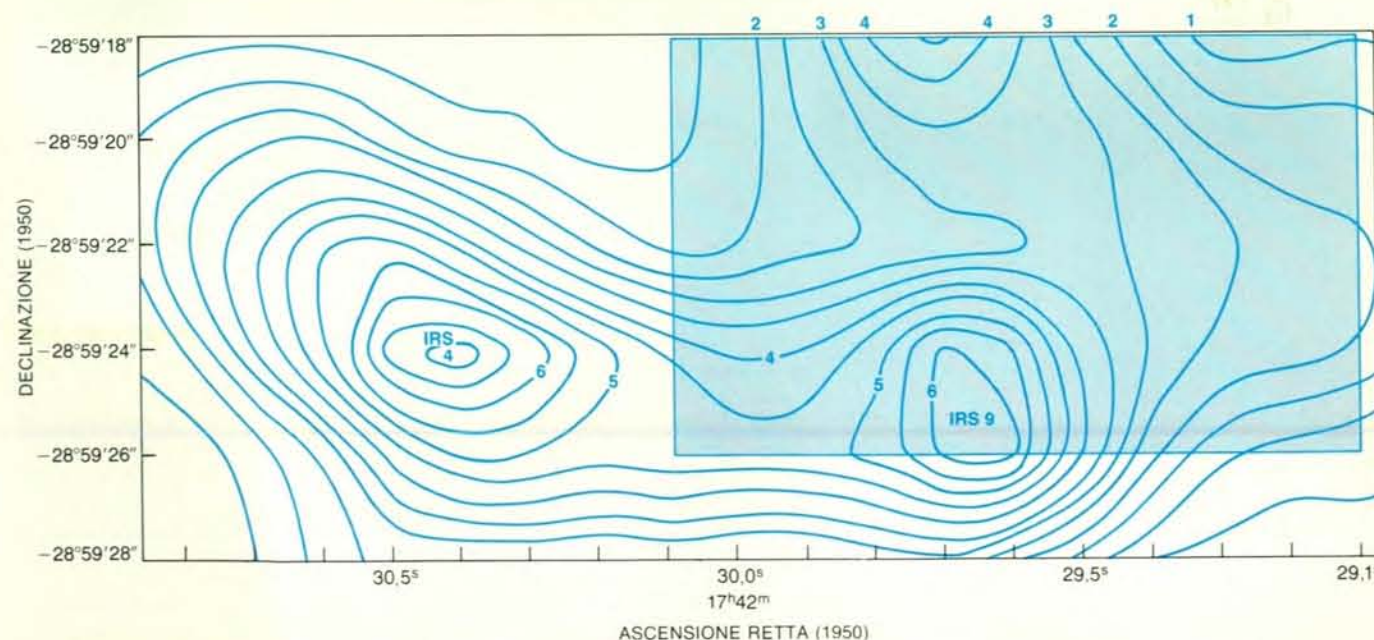
Data l'attrazione gravitazionale verso il centro della Galassia non è sorprendente la scoperta di un denso ammasso di stelle nel centro. Probabilmente l'ammasso si arricchisce di stelle ordinarie che orbitavano in precedenza a distanze maggiori dal centro galattico e che hanno perso momento angolare, forse a seguito di incontri ravvicinati con altre stelle. Viceversa, per lo stesso meccanismo, stelle del nucleo possono acquistare talvolta velocità sufficienti per fuggir via. Pertanto, la popolazione stellare del parsec centrale resta probabilmente in uno stato pressoché stazionario, oppure, al più, cambia molto lentamente.

Al contrario, concentrazioni dense di polvere e gas ionizzato hanno di solito vita breve. Nelle regioni esterne della Galassia ci sono due classi di questi oggetti: le regioni H II che si trovano in vicinanza di stelle calde e luminose di recente formazione, e le nebulose planetarie, che sono anch'esse regioni H II ma che devono il loro stato di ionizzazione a certe vecchie stelle più calde, ma meno luminose. In effetti, occorrerebbe la radiazione ultravioletta di circa 100 stelle analoghe a quelle che ionizzano le nebulose planetarie per spiegare la quantità di gas ionizzato osservata nel parsec centrale, ma, come abbiamo visto, sono state osservate solo circa 10 nubi ionizzate. Un'identificazione delle nubi con nebulose planetarie potrebbe essere messa in dubbio per questa ragione. D'altra parte, per avere lo stesso effetto basterebbero circa 10 stelle altamente luminose di recente formazione analoghe a quelle che ionizzano le normali regioni H II. Altre sorgenti tipiche di polvere e gas, come il «vento» di particelle emesso da stelle, o la caduta verso l'interno di materia dalle regioni circostanti non darebbero luogo alla distribuzione fortemente disuniforme della materia interstellare, che è stata dedotta dalle osservazioni del continuo infrarosso e della riga del neo.

In una regione densamente popolata di stelle come il parsec centrale bisogna considerare, però, un altro meccanismo per produrre nubi dense: le collisioni stellari. In una collisione distruttiva, in cui due stelle entrano effettivamente in contatto, verrà strappata materia da una stella o da entrambe, creando così una nube turbolenta e inizialmente densa che si disperde nell'ambiente circostante più rarefatto. Calcoli di Townes, Lacy e David Hollenbach indicano che nel parsec centrale le collisioni a cui prende parte una gigante rossa costituiscono una fonte importante di materia interstellare. Le nubi che si formerebbero a seguito di queste collisioni hanno dimensioni, velocità, moti interni e tempi di vita coerenti con le proprietà delle nubi osservate effettivamente. Inoltre, è possibile che il tasso di collisioni stimato sia adeguato a fornire nubi nel numero appropriato.

Il problema successivo è: qual è la sorgente di radiazione ionizzante necessaria per trasformare le nubi del parsec centrale in regioni H II? Ho indicato le stelle calde e luminose e le nebulose planetarie come possibili agenti ionizzanti. Le stelle luminose ad alta temperatura rimangono tali solo per pochi milioni di anni, e perciò non possono essersi formate in una regione molto lontana da quella in cui si osservano oggi, cioè nel nucleo galattico stesso. È opinione comune, però, che la formazione stellare sia un processo soggetto a un bilancio delicato che comporta la contrazione di grandi nubi di gas e polvere nel corso di circa un milione di anni. Nell'ambiente turbolento del nucleo galattico, in cui si hanno periodi orbitali di soli 10 000 anni, grandi volumi di gas e polvere non potrebbero contrarsi in stelle prima di essere distrutti completamente.

Le restrizioni più importanti sul tipo di sorgente ionizzante nel parsec centrale sono poste dal livello di ionizzazione uniformemente basso della regione. Le regioni H II normali, che si trovano in altre regioni della Galassia, contengono di solito, oltre al neo ionizzato una sola volta, anche specie atomiche a ionizzazione multipla, a cui mancano, cioè, due o più elettroni. Per creare gli stati di ionizzazione più elevata occorrono fotoni ultravioletti molto energetici provenienti da stelle estremamente calde.

Nel corso delle nostre osservazioni del Ne II nel parsec centrale abbiamo cercato attivamente anche righe spettrali dell'argo ionizzato due volte (Ar III) e dello zolfo ionizzato tre volte (S IV) in corrispondenza delle loro lunghezze d'onda infrarosse note. Spesso queste righe sono molto intense nelle regioni H II fuori dal nucleo galattico. I risultati in varie posizioni interne al parsec centrale sono stati negativi, con l'eccezione di una riga molto debole dell'Ar III rivelata in direzione di IRS 1. Un'intensa riga infrarossa dell'Ar II (argo ionizzato una sola volta) emessa dal parsec centrale è stata osservata recentemente da Steven P. Willner, Ray Russel e Richard Puetter dell'Università della California a San Diego, da Thomas Soifer del Caltech e da Paul Harvey dell'Università dell'Arizona. Il loro risultato comporta una moderata sovrabbondanza

Lo studio dettagliato della riga del neo mostra la complessità dei movimenti all'interno del parsec centrale della Galassia. L'insieme di spettri è disposto secondo la posizione nel cielo e comprende un po' più del parsec centrale. Un cerchio nell'angolo in basso a destra dà le dimensioni del fascio dello spettrometro; una scala dà la velocità di movimento del gas in chilometri al secondo. In ogni spettro l'origine coincide col valore zero della velocità lungo la linea di vista e dell'intensità dell'emissione, e con la posizione verso la quale era puntato il telescopio. Numerose componenti di velocità localizzate della riga del neo coincidono con sorgenti che sono brillanti a 10 micrometri (IRS 1, IRS 2, IRS 6, IRS 9 e IRS 10), quindi il gas ionizzato è associato fisicamente alle nubi di polvere calde. La linea tratteggiata separa regioni in cui predomina neo ionizzato in avvicinamento *(a destra)* da quelle in cui predomina neo ionizzato in allontanamento *(a sinistra)*, indicando che la materia nel parsec centrale ruota nel senso dato dalla freccia. La regione contenuta nel rettangolo è riprodotta sotto.

L'intensità della riga del neo nella parte sudorientale del parsec centrale dimostra la associazione del neo ionizzato con due particolari sorgenti, IRS 4 e IRS 9, la cui radiazione continua è prodotta in nubi di polvere calde. Le curve di livello mostrano la intensità del neo con velocità pari a circa +125 chilometri al secondo. I picchi coincidono con le due sorgenti infrarosse. Spettri misurati alle posizioni delle sorgenti indicano che in effetti IRS 4 si allontana alla velocità di 110 chilometri al secondo e IRS 9 a 140 chilometri al secondo. Le osservazioni per le illustrazioni di questa pagina sono state eseguite dall'autore e dai suoi colleghi col telescopio da 2,5 metri dell'Osservatorio di Las Campañas.

di argo nel parsec centrale e indica chiaramente, una volta combinato coi nostri risultati, che lo stato di ionizzazione nel parsec centrale è insolitamente basso.

Le stelle centrali delle nebulose planetarie, che di solito hanno temperature molto elevate, sembrerebbero escluse come sorgenti di radiazione ionizzante, dato che per la maggior parte generano stati di ionizzazione elevata. La situazione resta un po' singolare, però, anche se la sorgente principale è costituita da molte stelle luminose ad alta temperatura di recente formazione. Infatti, ci si aspetterebbe che le stelle ionizzanti presentino un ampio intervallo di temperature, a meno che la formazione delle stelle più calde e massicce risulti in qualche modo impedita nel centro galattico, o a meno che quelle di più recente formazione si siano raffreddate significativamente. Questa varietà di temperature porterebbe a variazioni dello stato di ionizzazione attraverso il parsec centrale. L'assenza di Ar III e S IV in tutto il parsec centrale sembra indicare o che tutte le stelle ionizzanti sono pressoché identiche, oppure che nel parsec centrale c'è una sola sorgente ionizzante estremamente luminosa, ma a temperatura relativamente bassa. Un oggetto di quest'ultimo tipo non potrebbe costituire una stella ordinaria.

Queste considerazioni portano a un modello del parsec centrale piuttosto particolare: un modello caratterizzato dalla presenza di un buco nero massiccio nel centro. Si pensa che il materiale attirato nel buco nero sia immagazzinato dapprima in un disco in rotazione, che in condizioni appropriate irradierebbe intensamente nell'ultravioletto. La presenza di un oggetto di questo tipo potrebbe spiegare l'uniformità dello stato di ionizzazione del parsec centrale. Un candidato naturale per il buco nero è la radiosorgente compatta, di diametro inferiore a 0,001 secondi d'arco, in Sagittario *A* Ovest.

Le proprietà previste per i buchi neri massicci sono state calcolate da molti astrofisici teorici. Lo spettro e l'intensità della radiazione ionizzante emessa dal disco dipendono solo dalla massa e dal tasso di accrescimento del buco nero. Townes, Lacy e Hollenbach hanno mostrato che ci sarebbe un accordo quantitativo con la luminosità e lo stato di ionizzazione osservati nelle regioni H II del parsec centrale della Galassia ipotizzando la presenza di un buco nero che avesse una massa pari a qualche milione di masse solari e che acquistasse materia al tasso relativamente modesto, di 10^{-5} masse solari per anno. Una massa molto superiore a quella necessaria per mantenere questo tasso di accrescimento potrebbe essere fornita da caduta di materia, venti stellari, nebulose planetarie e dai resti di collisioni stellari nel parsec centrale. Anche se si ammette che parte del materiale sia soffiato via dalla intensa pressione della radiazione emessa dal disco di accrescimento, da circa due milioni di stelle e da altri oggetti luminosi del parsec centrale, sembrerebbe che resti abbastanza materiale per alimentare ampiamente il buco nero al tasso previsto.

Forse è significativo il fatto che l'intensità della radiazione infrarossa emessa dall'oggetto peculiare IRS 16, nel quale sembra sia localizzata la radiosorgente compatta, è circa uguale all'intensità che sarebbe emessa in prossimità di un buco nero che emetta materia al tasso sopraindicato. Infine, l'esistenza di un buco nero con una massa pari a qualche milione di masse solari è senz'altro coerente con la differenza tra la massa totale stimata nel parsec centrale e la massa delle stelle contenute nello stesso volume. Un buco nero di questa massa sta comodamente in un volume pari a quello della radiosorgente compatta.

Le prove a favore della presenza di un buco nero massiccio sembrano ragionevoli, ma sono indirette. Prove più dirette potrebbero essere raccolte in future osservazioni radio e infrarosse della piccola regione attorno alle radiosorgenti compatte. Per esempio, l'intenso campo gravitazionale in vicinanza di un buco nero di tale massa dovrebbe determinare un allargamento Doppler di 1000 o più chilometri al secondo nelle righe spettrali che si generano nella sorgente. Quando il nostro spettrometro infrarosso, con un campo di 0,2 parsec di diametro alla distanza del centro galattico, era puntato sulla radiosorgente compatta (una possibile localizzazione del buco nero) la riga del neo non risultava così larga. Ciononostante, poteva essere presente una componente ampia e debole che il nostro strumento non riusciva a rivelare. Cambiamenti significativi nel tasso di accrescimento del buco nero, che possono aver luogo in tempi dell'ordine di qualche anno soltanto, finirebbero per determinare variazioni osservabili dello stato di ionizzazione del parsec centrale. Pertanto è importante tenere sotto continuo controllo tutte le righe di emissione infrarosse.

Ci si potrebbe chiedere: stiamo osservando davvero il parsec centrale della Galassia? Secondo me la risposta è decisamente sì. Prima di tutto, stiamo certamente puntando i nostri telescopi nella direzione del parsec centrale; questo fatto è indicato da tutte le osservazioni infrarosse e radio, e forse nel modo più chiaro dalle osservazioni della distribuzione stellare condotte a 2,2 micrometri. Secondo, in nessun'altra parte della Galassia si osservano così tante sorgenti infrarosse. Le alte velocità positive e negative ottenute dalla riga del neo sono uniche, e le enormi densità di massa indicate dalle misure a 2,2 micrometri e da quelle della riga del neo non sono state riscontrate da nessun'altra parte.

Il parsec centrale può essere l'origine di fenomeni che si osservano ben al di fuori del nucleo galattico. Grandi aumenti improvvisi del tasso di accrescimento del buco nero proposto, oppure supernove o esplosioni di formazione stellare presso il centro galattico potrebbero spiegare le nubi molecolari in moto verso l'esterno, che sono state osservate dai radioastronomi. Queste emissioni esplosive espellerebbero materia dal nucleo galattico sotto forma di nubi ad alta velocità. Il materiale finirà per entrare in collisione con le nubi molecolari osservate accelerandole; queste nubi potrebbero essere a loro volta i resti di esplosioni precedenti.

Nonostante che il nucleo della nostra galassia sia una potente sorgente di energia, è molto debole rispetto a molti altri nuclei galattici. Forse questa è una delle ragioni per cui noi esistiamo e possiamo studiare il nucleo galattico. Attualmente potrebbe attraversare una fase quieta tra epoche di attività più esplosiva. Lo studio continuato del nucleo galattico ci consentirà una migliore comprensione delle sue proprietà e forse anche degli eventi fisici che hanno sede nei nuclei di galassie più attive.

Visione schematica del nucleo galattico, esteso per circa tre parsec, nell'ipotesi dell'esistenza di un buco nero massiccio (rappresentato molto più in grande del resto) proprio nel centro. Le aree in colore rappresentano nubi di gas e polvere interstellare che sono create al ritmo di una ogni 1000 anni circa dalle collisioni stellari che avvengono nel denso ammasso centrale di stelle. Una parte della polvere e del gas, attirata gravitazionalmente verso il buco nero, forma un disco di accrescimento caldo rotante prima di essere inghiottita e scomparire. L'intensa emissione di radiazione ultravioletta dal disco di accrescimento ionizza il gas delle nubi circostanti, che emette poi le righe infrarosse e il continuo radio, che sono stati osservati nella direzione del parsec centrale. Anche la polvere delle nubi è riscaldata dalla radiazione proveniente dal disco di accrescimento e dalle stelle vicine, ed emette anch'essa la radiazione continua infrarossa. Gran parte della radiazione infrarossa a lunghezze d'onda corte viene emessa da alcune giganti rosse *(cerchi in colore)* che sono localizzate nel denso ammasso centrale di stelle. I punti bianchi nel disegno rappresentano stelle normali, che sono valutabili intorno ai due milioni nel parsec centrale.

Le origini del primo aeroplano a motore pilotato

Nel 1903 il «Flyer» dei famosi costruttori di biciclette, i fratelli Wright, non fu un tentativo più fortunato di altri, ma il risultato di un intenso programma di ricerche teoriche e pratiche e di numerose prove

di F. E. C. Culick

Wilbur e Orville Wright, con una serie di voli il 17 dicembre 1903, furono i primi uomini a pilotare un aeroplano a motore. Come si sa bene, essi stessi avevano progettato e costruito l'apparecchio. E tuttavia ancora oggi si continua a credere che i Wright fossero soltanto due meccanici di biciclette inciampati più o meno per caso in un progetto di successo. Anche tra gli ingegneri aeronautici è scarsa la conoscenza di quello che i Wright hanno fatto. Ricostruendo i passi attraverso i quali i fratelli arrivarono alla loro vittoria e costruendo un modello del loro «Flyer» del 1903 per provarlo in una galleria del vento, sono arrivato a capire di quale straordinaria qualità sia stato ciò che hanno fatto nella ricerca, nell'ingegneria e nella sperimentazione. Il loro lavoro merita di essere inquadrato in un più ricco contesto storico.

Il giorno dei loro primi voli i fratelli si trasformarono in piloti. Wilbur era il pilota del quarto volo, l'ultimo e il più lungo della giornata: ben 59 secondi. Dovevano passare quasi quattro anni prima che chiunque altro fosse capace di volare per un minuto, e anche allora la macchina non sarebbe stata completamente controllabile. E in quel frattempo i Wright avevano messo a punto un altro aeroplano manovrabile che era in grado di volare per più di un'ora.

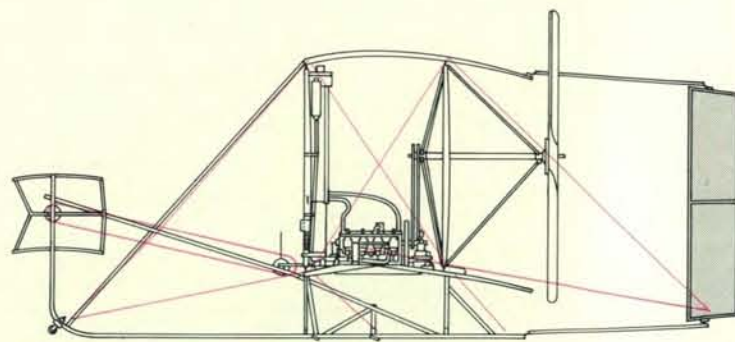
Nel 1908, quando i Wright volarono per la prima volta in pubblico, nessun altro aveva ancora capito la necessità del controllo laterale, molto meno la funzione del timone verticale. Perciò nessun altro poteva fare una virata vera. Né alcun altro sapeva come costruire delle buone eliche. Soprattutto, però, nessun altro aveva mai portato avanti un programma come il loro: conducendo le necessarie ricerche, costruendo il proprio aereo e facendolo volare, così da padroneggiare l'intero problema. I Wright avevano potuto superare tutti gli altri che tentavano di volare perché essi avevano individuato i problemi che dovevano essere affrontati e risolti.

I fratelli cominciarono il loro lavoro nel 1899, quando Wilbur aveva 32 anni e Orville 28. Wilbur, naturale guida per molta parte del lavoro, si era reso conto che si sarebbe dovuto seguire un programma sistematico di ricerca e di ingegneria. Sintetizzò il suo punto di vista in una lettura che tenne nel settembre 1901 a una riunione della Western Society of Engineers: «I problemi del viaggio terrestre e marittimo furono risolti nel diciannovesimo secolo perché era stato possibile incominciare con piccoli passi per poi allargare gradualmente il lavoro sino alla presente situazione di sviluppo. Il problema del volo fu invece lasciato al ventesimo secolo perché prima di poter compiere un volo di durata significativa bisognava che le conoscenze progredissero molto». Quattro anni dopo i Wright avrebbero avuto un aeroplano.

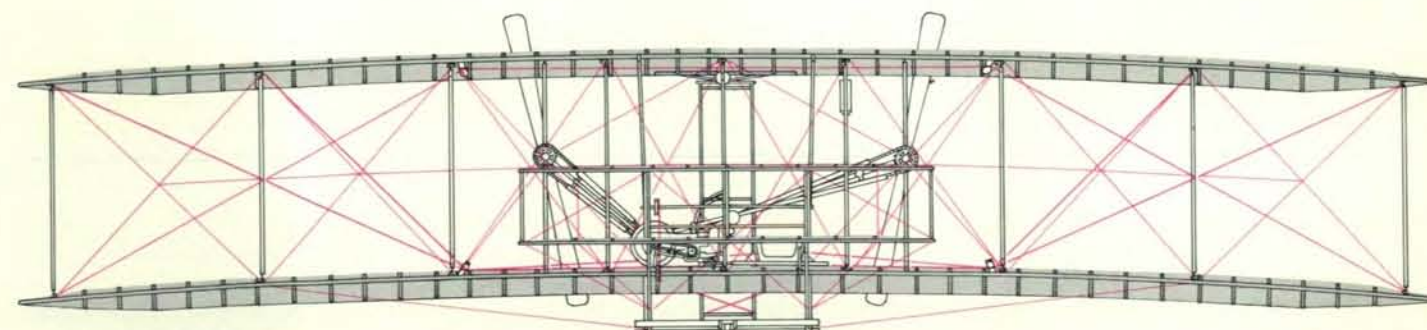
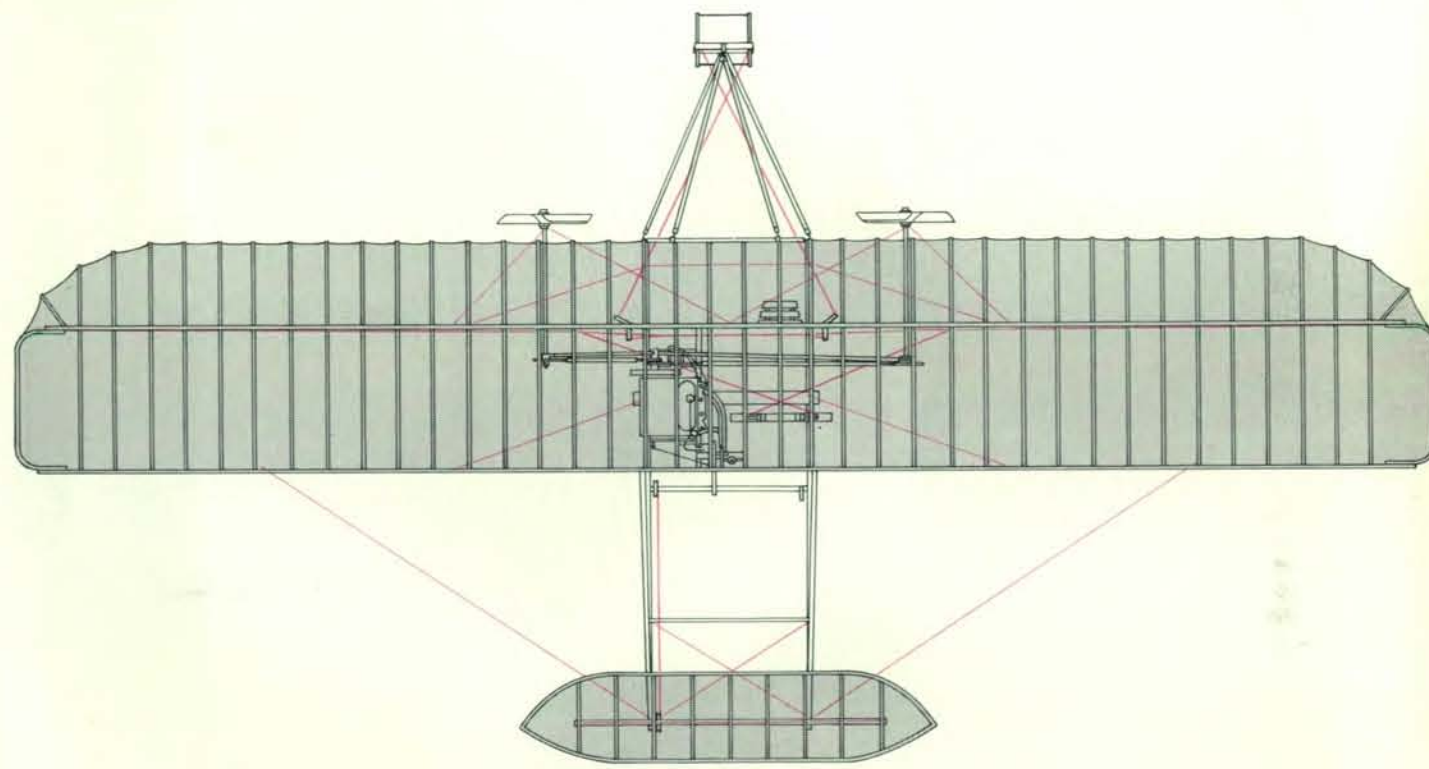
La loro invenzione dell'aeroplano a motore capace di portare uomini a bordo nacque in parte come naturale prosecuzione del lavoro di altri. Tuttavia il contributo peculiare dei Wright fu senz'altro determinante. I risultati tecnici che riuscirono a conseguire erano sbalorditivi, specialmente se si considera a che punto erano nel 1899 gli sforzi per arrivare al volo umano.

Quattro pionieri

Coloro che cercavano di costruire macchine volanti avevano seguito diverse strategie. Le osservazioni del volo degli uccelli stimolarono gli studi, e così naturalmente le prime idee di soluzione del problema comportavano sempre l'uso di ali battenti. Sir George Cayley (nel 1799) fu il primo a capire che il volo umano sarebbe stato realizzato assai più agevolmente se il modo di generare la portanza fosse stato realizzato separatamente dai mezzi di propulsione.

Cayley è riconosciuto come il padre dell'aeroplano. I suoi progetti e le sue conclusioni si fondavano su osservazioni accurate e su esperimenti. Fu lui a concepire la configurazione oggi definita convenzionale: un corpo o fusoliera che porta un'ala principale, con impennaggi orizzontali e verticali disposti a poppa. Fece volare con successo alcuni veleggiatori, uno dei quali grande abbastanza da portare un ragazzo, almeno una volta. Cayley non volò mai. Fra gli altri suoi notevoli contributi al volo ci sono la dimostrazione che le superfici curve sono migliori di quelle piane per generare portanza; l'idea che la stabilità trasversale può essere ottenuta dando all'ala un angolo diedro, ovvero costruendola con le estremità rialzate; infine, il concetto dell'utilizzazione di un timone orizzontale per ottenere la stabilità longitudinale.

Quello che Cayley incominciò a capire fu lo stretto rapporto tra la geometria di un aereo, le forze che agiscono su di esso e la sua stabilità in volo. Un aeroplano ha un piano verticale di simmetria che passa attraverso il suo asse longitudinale. I moti che non deviano il velivolo in una direzione fuori da questo piano sono chiamati moti longitudinali. Il moto più comune è il beccheggio, nel quale la prua si muove in su e in giù. Il rollio intorno all'asse longitudinale e l'imbardata intorno all'asse verticale sono globalmente chiamati moti laterali. La tendenza di ogni oggetto in movimento continuo a ritornare al suo stato iniziale in seguito a una perturbazione si chiama stabilità. Ci sono due tipi di stabilità, longitudinale e laterale, corrispondenti a due classi di moti dell'aeroplano. In un moderno aereo la stabilità longitudinale è ottenuta soprattutto a mezzo di una piccola superficie orizzontale nella parte posteriore, secondo le idee di Cayley. La stabilità laterale è effetto di molte cause, compreso l'angolo diedro delle ali e l'impennaggio verticale. Il disegno delle prime macchine volanti comprendeva un impennaggio verticale per la guida, in analogia con il timone di una nave.

Il gigante della prima aviazione francese fu Alphonse Pénaud. Fu il primo a usare una treccia di elastico per far girare l'elica di un modello d'aeroplano. Indipendentemente da Cayley pensò all'impennaggio orizzontale come un mezzo per ottenere la stabilità longitudinale. Di più, capì come «lavora» l'impennaggio e diede la prima spiegazione della sua funzione nel determinare un volo stabile.

Otto Lilienthal, un ingegnere meccanico che lavorava in Germania, fu quello che più direttamente influenzò i Wright.

Il «Flyer» dei Wright del 1903, a sinistra visto dal lato sinistro e, a destra, dall'alto e di fronte, è stato il primo aereo a motore pilotato che abbia volato con successo. La sua caratteristica più evidente era il timone anteriore orizzontale. Il doppio timone verticale posteriore non si muoveva, salvo quando il pilota svergolava l'ala per compiere una virata. Volando in posizione prona, il pilota otteneva lo svergolamento a mezzo di cavi agganciati a una specie di culla sulla quale si trovava. Lo svergolamento alare, un'invenzione dei Wright, assolveva alle funzioni oggi proprie degli alettoni. Wilbur e Orville Wright costruirono tutto l'aeroplano, compreso il motore a benzina e le eliche che avevano progettato da zero.

Nel 1889 pubblicò un libro *Il volo degli uccelli come base per l'arte dell'aviazione,* che conteneva i risultati di una estesa serie di esperimenti condotti con l'aiuto del fratello. Il lavoro presentava i primi dati utili sulla portanza e sulla resistenza aerodinamica di ali curve. Lilienthal si provò anche nel volo planato, a partire dal 1891, e divenne il primo pilota a riuscire nel volo a vela. Sperimentò alianti monoplani e biplani e rimase ucciso mentre provava un monoplano.

In questo stesso periodo, negli Stati Uniti il più serio studioso di aeronautica fu Octave Chanute, un eminente ingegnere civile. Con i suoi scritti e con i suoi esperimenti fece conoscere nel suo paese il lavoro di Lilienthal. L'età (aveva 68 anni nel 1900) gli impedì di tentare il volo di persona. Ma aveva diversi discepoli che sperimentarono i suoi alianti insieme con i propri progetti.

Chanute progettò un aliante biplano che rappresenta un po' il capostipite dell'aeroplano moderno. (Il progetto di recente è stato preso come base per diversi libratori, alianti che possono soltanto scendere.) Chanute adottò l'impennaggio posteriore di Pénaud per la stabilità longitudinale, mentre per ottenere una maggiore portanza progettò una sezione alare curva simile a una provata da Lilienthal.

Meno noto è il più importante contributo di Chanute alla tecnica aeronautica: la sua scelta della struttura per il biplano. Adottò la «travatura» di Pratt che era stata brevettata nel 1844 per i ponti ferroviari. Le due ali erano collegate da solidi montanti verticali che sopportavano i carichi di compressione. I carichi di tensione erano trasmessi da cavi diagonali (tiranti) incrociati che si agganciavano ai montanti sia nel senso dell'apertura alare sia tra parte anteriore e parte posteriore dell'ala. Ne risultava una struttura rigida, ma leggera che divenne la base per tutti i biplani controventati. (Si definiscono controventate le ali con montanti e tiranti esterni, mentre quelle senza strutture esterne si definiscono a sbalzo.)

Lilienthal e Chanute cercavano di risolvere il problema della stabilità longitudinale. Erano capaci di dare ai loro veleggiatori una portanza sufficiente, ma trovavano gravi difficoltà a mantenere l'equilibrio in volo. Capirono che ottenere l'equilibrio significa far coincidere il centro di spinta (il punto dove si equilibrano le forze aerodinamiche che agiscono sul velivolo) con il centro di gravità. Il guaio è che il centro di spinta si sposta se l'assetto dell'aeroplano viene modificato, per esempio da una raffica di vento. In un libratore il pilota può ripristinare l'equilibrio spostandosi un poco per variare il centro di gravità e cambiare l'assetto del velivolo. Se questa manovra è compiuta correttamente, la combinazione dell'apparecchio e del pilota forma un sistema stabile.

Quando un impennaggio orizzontale posteriore è sistemato nella maniera giusta, la sua azione (come Pénaud aveva capito) sarà tale da riportare il velivolo in equilibrio. La ragione è che l'effettivo centro di spinta di ala e impennaggio, considerati insieme, resta fisso anche se l'assetto della macchina viene variato. L'aeroplano ha una sua stabilità intrinseca, e il pilota non ha bisogno di cambiare la sua posizione. Un modello di aereo è stabile proprio per questo. Un aeroplano vero richiede più controllo, e per questo l'impennaggio è mobile, e il pilota può cambiarne l'azione a volontà. Con un impennaggio mobile un aeroplano instabile può essere guidato da un pilota sufficientemente esperto. (Quando l'impennaggio diventa mobile, in realtà è solo una parte di esso che si muove. Così si definisce impennaggio il tutto, e timone la parte mobile. La più moderna tecnica aeronautica, è tornata a impennaggi praticamente in un pezzo solo, che però sono mobili, e cioè sono tutto timone.)

I fratelli Wright cominciarono a pensa-

Il primo veleggiatore fu progettato da Otto Lilienthal, che si vede qui mentre vola sopra una fabbrica di mattoni nei sobborghi di Berlino, nel 1894. Lilienthal fu il primo pilota di quello che adesso sarebbe chiamato un libratore, cioè un veleggiatore che può soltanto planare, scendere. Rimase ucciso nel 1896 quando uno dei suoi alianti andò in stallo e precipitò, appunto per la difficoltà di mantenere l'equilibrio in volo. La linea orizzontale irregolare che passa vicino alle teste degli spettatori in questa fotografia è una rottura della negativa originale, di vetro.

Un aliante biplano fu progettato nel 1896 da Octave Chanute, un ingegnere civile che fu il più serio studioso di aeronautica negli Stati Uniti prima che i Wright incominciassero il loro lavoro. Ebbe con essi una lunga corrispondenza, fu consultato per molti dei loro primi voli. Il pilota impegnato in questo volo del 1897 nell'Indiana fu A. M. Herring, uno degli assistenti di Chanute.

re al volo quando, ragazzini, ebbero in dono un elicottero-giocattolo, simile a uno inventato da Pénaud. Più tardi, nel 1896, le notizie sulla morte di Lilienthal stimolarono il loro interesse per il problema del volo. Nel maggio del 1899, Wilbur scrisse al segretario della Smithsonian Institution chiedendo un elenco delle opere disponibili sul volo. Di quella lista decise di acquistare un libro di Chanute, *Progress in Flying Machines* (Progresso nelle macchine volanti) scritto nel 1894. Dopo aver studiato il libro, Wilbur scrisse a Chanute, aprendo una corrispondenza che durò quasi dieci anni. Per buona parte di quel periodo, Chanute fu una fonte di informazioni e di incoraggiamento per i Wright. Le lettere e i diari dei Wright costituiscono una cronaca particolareggiata dalla quale si possono ricostruire gli eventi che portano al volo a motore.

L'inizio dei Wright

Mentre cominciavano una seria sperimentazione, i fratelli si erano familiarizzati con la letteratura aeronautica. Si erano resi conto che secondo le opinioni prevalenti la stabilità longitudinale richiedeva un impennaggio alla Pénaud, che la stabilità laterale poteva essere ottenuta dando alle ali un angolo diedro e che le variazioni di direzione si potevano attuare mediante un impennaggio verticale o timone. Ma essi non accettarono nessuno di questi tre principi, preparando progetti completamente diversi da quello base di Cayley, generalmente accettato dai loro contemporanei. Il loro primo aeroplano a motore, il Flyer del 1903, aveva una superficie orizzontale anteriore, le estremità alari abbassate in un diedro negativo e l'impennaggio verticale che si muoveva soltanto quando le ali venivano flesse per far girare l'aeroplano. Le ragioni per le quali i fratelli Wright adottarono questa configurazione raccontano la storia delle loro ricerche e delle loro soluzioni.

Probabilmente la più importante motivazione nel programma dei Wright era che essi erano decisi a imparare a volare. Questa determinazione influenzò praticamente ogni loro decisione. Poiché dovevano diventare piloti essi dovevano considerare con molta attenzione i problemi della stabilità, del controllo e della sicurezza. Non avevano intenzione di costruire una macchina nella speranza che funzionasse e poi «noleggiare» qualcun altro per fargliela collaudare. E chiaramente non erano per nulla soddisfatti di scendere veleggiando diritto davanti a loro, e basta.

La morte di Lilienthal e un altro incidente del genere in Inghilterra convinsero Wilbur fin da principio che, per essere sicuro, un aeroplano avrebbe dovuto avere una superficie orizzontale stabilizzatrice davanti all'ala principale (un sistema che sarebbe stato chiamato *canard,* anatra). Nessuno a quel tempo conosceva lo stallo (la perdita di portanza), né i Wright avevano basi scientifiche per stabilire i rispettivi vantaggi delle superfici anteriori e posteriori. Mentre essi conducevano esperimenti, le condizioni teoriche per ottenere la stabilità venivano sviluppate da altri, ma fino a dopo il 1903 non si seppe che un aereo poteva ottenere la stabilità longitudinale con una superficie orizzontale sia anteriore sia posteriore. Il piano anteriore, che era il particolare più insolito dei primi aeroplani Wright, aveva il vantaggio, per i fratelli, che il pilota lo poteva vedere, guadagnando così in sicurezza perché poteva reagire rapidamente se non funzionava bene.

Contrariamente ai suoi predecessori, Wilbur non dette per scontato che il problema della stabilità laterale fosse già stato risolto dall'invenzione, di Cayley, del diedro alare. «Le mie osservazioni sul volo delle poiane - scrisse Wilbur a Chanute - mi portarono a pensare che esse ripristinano il loro equilibrio laterale, quando viene turbato da un colpo di vento, con una torsione delle estremità alari.» Fatta questa fondamentale osservazione, Wilbur doveva trovare un metodo analogo per il controllo di un aereo con un uomo a bordo.

Trovò la soluzione con lo svergolamento alare (mentre l'estremità destra viene sollevata, l'estremità sinistra viene abbassata, o viceversa), un sistema che ha preceduto gli alettoni, i quali assolvono alla stessa funzione negli aeroplani moderni. Con l'ala svergolata i Wright non soltanto riuscivano a mantenere stabile l'aereo in volo orizzontale ma anche a farlo virare. Così essi andarono molto più avanti dei loro concorrenti. Utilizzando una superficie orizzontale mobile e l'ala svergolata, i fratelli Wright riunirono stabilità e controllo, apportando con questo il loro più importante contributo allo sviluppo dell'aeroplano moderno.

Parecchi anni dopo che Wilbur aveva scritto alla Smithsonian Institution, costruì un aquilone biplano con un'apertura alare di un metro e mezzo. Gli aquiloni come ora li conosciamo (in particolare gli aquiloni a scatola) sono stati sviluppati piuttosto di recente. Wilbur seguì le ultime regole di costruzione, ma la struttura base del suo aquilone era la «travatura» di Pratt, che aveva preso da Chanute. Tolse i tiranti di rinforzo avanti-dietro per avere le ali libere di svergolarsi una rispetto all'altra. Le sole parti rigide della struttura erano le connessione tra la superficie di controllo e il montante al quale era unita. I quattro cavi di comando erano sistemati in modo che il pilota potesse avere il controllo sia longitudinale sia laterale dell'aquilone, contemporaneamente o separatamente.

Prove in volo

Wilbur mandò in volo la sua prima macchina sperimentale un pomeriggio, nell'agosto del 1899, mentre Orville era in vacanza. Sebbene gli aquiloni del genere fossero piuttosto sensibili in aria, Wilbur riuscì a mantenerlo in volo, convincendosi che era valida la sua idea per il controllo longitudinale e laterale. Trovò

anche che l'aquilone poteva volare con il timone orizzontale montato davanti o indifferentemente dietro. Quel notevole aquilone fu la prima macchina volante che poteva essere controllata lateralmente e longitudinalmente.

Il progetto dell'aquilone del 1899 divenne la base per una versione ingrandita, abbastanza piccola da poter volare come un aquilone, ma abbastanza grande da poter portare un uomo. Aveva un'apertura alare di cinque metri e 10 centimetri con una corda alare (larghezza) di un metro e mezzo; il suo peso a vuoto era poco più di 22 chili, 86 chili con il pilota che si metteva in posizione prona. Con il pilota a bordo il carico alare era di 560 grammi per metro quadrato. (Un veleggiatore moderno di avanzate caratteristiche può avere un carico alare di oltre 24 chilogrammi per metro quadrato.

Wilbur pensava di far volare questo suo velivolo nel 1900, in un primo tempo facendolo partire da una torre, trattenuto da cavi fissati alla torre stessa. Se i venti fossero stati costanti, lui e il fratello avrebbero potuto accumulare ore di esperienza manovrando i comandi senza i rischi del volo libero. Essi avevano calcolato che in cinque anni Lilienthal aveva passato soltanto cinque ore in aria, che a malapena è il tempo necessario per imparare a volare.

Chanute suggerì un certo numero di luoghi dove i Wright avrebbero potuto trovare venti costanti nella stagione durante la quale potevano fare i loro esperimenti (a causa degli impegni derivanti dalla loro fabbrica di biciclette potevano lasciare la casa di Dayton, nell'Ohio, soltanto fra settembre e gennaio). Dopo aver verificato con l'Ufficio meteorologico le probabili condizioni del vento in varie località, decisero di andare a Kitty Hawk, sulla costa della Carolina del Nord. Kitty Hawk era un piccolo abitato, isolato, che si poteva raggiungere soltanto dal mare dopo un viaggio in treno dall'Ohio. I due fratelli non soltanto dovevano trasportare l'aereo e tutto l'equipaggiamento dalla terraferma ma anche trovare il materiale per costruire la torre.

Nel 1900 i Wright rimasero meno di un mese presso Kitty Hawk, provando il loro aliante dal 5 al 18 ottobre. Per la maggior parte del tempo fecero volare l'aereo come un aquilone, prima da una torre alta tre metri poi da terra con cavi tenuti a mano. In tutto, i voli liberi totalizzarono due minuti, dei quali approssimativamente 10 secondi con un uomo a bordo. Soltanto Wilbur volò. I comandi si comportavano bene quando l'aliante volava come aquilone, mentre in quei brevissimi 10 secondi di volo pilotato Wilbur non fu in grado di padroneggiare l'adattamento dei comandi a «guidare» le manovre.

Risulta chiaro dai diari che i primi esperimenti furono spesso deludenti. Orville annotava: «Lo provammo con l'impennaggio davanti, di dietro, e in ogni altro modo. Quando finimmo, Wil era così confuso da non riuscire a ricavare una teoria. E dovetti mettercela tutta per non fargli abbandonare la nostra attività di volo». I voli furono brevi e ci furono parecchi atterraggi duri. Fortunatamente la struttura era semplice e facile da riparare.

Nastri colorati attaccati ai cavi di trattenuta consentivano ai fratelli di misurare la portanza e la resistenza. Determinavano la velocità del vento con un anemometro. Quelli che raccoglievano erano i primi dati sulle caratteristiche di una struttura portante (nel senso aerodinamico) che si raccogliessero in quantità significativa.

Per quanto ridotto fosse, il programma di voli dei Wright nel 1900 fruttò risultati fondamentali e servì a «confermare la giustezza delle (loro) prime valutazioni». I fratelli verificarono le loro tecniche per ottenere la stabilità longitudinale e laterale: e ancor più decisamente stabilirono che dovevano addestrarsi al volo. E dai loro rilevamenti si resero conto che volare in posizione prona riduceva la resistenza della macchina.

Risultati delle prove

Non tutte le loro conclusioni, però, furono così positive. Fu allora che Wilbur decise di non costruire le ali con l'angolo diedro. Sebbene quella configurazione contribuisse alla stabilità in aria calma, le estremità leggermente rialzate delle ali tendevano ad accentuare l'effetto dei colpi di vento laterali. Poiché i Wright ottenevano il controllo laterale (trasversale) a mezzo dello svergolamento alare, non avevano più bisogno dell'effetto stabilizzatore del diedro alare.

I risultati più deludenti degli esperimenti del 1900 furono i valori trovati per la portanza e per la resistenza. La resistenza della macchina quando non si generava portanza era assai minore rispetto al previsto. Invece per un dato angolo tra l'ala e il vento la resistenza totale era maggiore e la portanza minore di quanto ci si aspettasse (i Wright avevano preso come base i dati rilevati da John Smeaton nel 1759 su una superficie piana perpendicolare a un flusso d'aria e sulle misure per un'ala curva fatte da Lilienthal). I fratelli osservarono nei loro voli che l'aliante procedeva con un angolo di incidenza parecchio superiore a quello da essi previsto. La portanza aumenta con l'angolo di incidenza e deve uguagliare il peso dell'aliante. Ma a un dato angolo di attacco le ali dei Wright producevano meno portanza di quanto i fratelli avevano previsto sulla base dei dati di Lilienthal.

Questo fu un risultato sorprendente, ma i Wright credevano nel loro lavoro e considerarono tre possibili spiegazioni: 1) il tessuto che ricopriva la struttura alare non era sufficientemente impermeabile all'aria; 2) la curvatura delle loro ali (1:22) era troppo poco accentuata perché fossero applicabili i dati di Lilienthal; 3) i dati di Lilienthal erano sbagliati. Essi propendevano per il secondo motivo e disegnarono l'ala del loro nuovo aliante, che doveva volare nell'autunno del 1901, con una curvatura più pronunciata.

Il primo aereo costruito dai Wright fu questo aquilone biplano del 1899. Aveva un'apertura alare di un metro e mezzo e utilizzava la «travatura» di Pratt disegnata per i ponti ferroviari e adattata per gli aerei da Chanute. L'aquilone incorporava anche il concetto di ala svergolata di Wilbur Wright. Le quattro corde consentono all'operatore il controllo trasversale e longitudinale.

L'aquilone-veleggiatore del 1900 fu il primo velivolo dei Wright abbastanza grande da portare un uomo. Era una versione ingrandita dell'aquilone del 1899, con un'apertura alare di 5 metri e 10 e una corda (profondità alare) di un metro e mezzo. I Wright lo provarono, soprattutto come aquilone, a Kitty Hawk, North Carolina, nell'ottobre del 1900. Soltanto Wilbur volò per un totale di circa dieci secondi. Sebbene i comandi funzionassero bene quando i Wright facevano volare il velivolo come aquilone, provando prima da una torre e poi da terra, Wilbur trovò nella sua breve esperienza di pilotaggio che non riusciva a padroneggiare l'adattamento dei comandi per «guidare» le manovre.

Con l'interpretazione dei risultati ottenuti nel 1900, i Wright non ebbero più necessità di «dipendere» dal lavoro di altri. Da quel momento i loro progressi derivarono esclusivamente dalle loro stesse ricerche. Cominciarono a lavorare durante l'inverno sull'aliante del 1901 (sempre un biplano). Per dargli una maggiore portanza aumentarono l'apertura alare a sei metri e sessanta. Con una corda di due metri e dieci, l'area totale alare ricoperta era di circa 27 metri quadrati. La superficie dello stabilizzatore era di poco più di un metro quadrato e mezzo. Era di gran lunga il più grosso aliante che si fosse mai tentato di far volare. Il peso della struttura era arrivato a 45 chili e il carico alare, con pilota a bordo, era di 380 grammi per metro quadrato. Le sole altre importanti modifiche furono l'aumento della curvatura del profilo alare da 1:22 a 1:12 e il cambio dei comandi così che il pilota potesse più agevolmente controllare lo stabilizzatore e lo svergolamento alare. Sebbene l'aliante non apparisse molto diverso da quello del 1900, per diversi aspetti esso risultava il più importante strumento di ricerca dei Wright. E in realtà, il 1901 fu l'anno decisivo per quasi tutti i loro successivi risultati.

Le prove si allargano

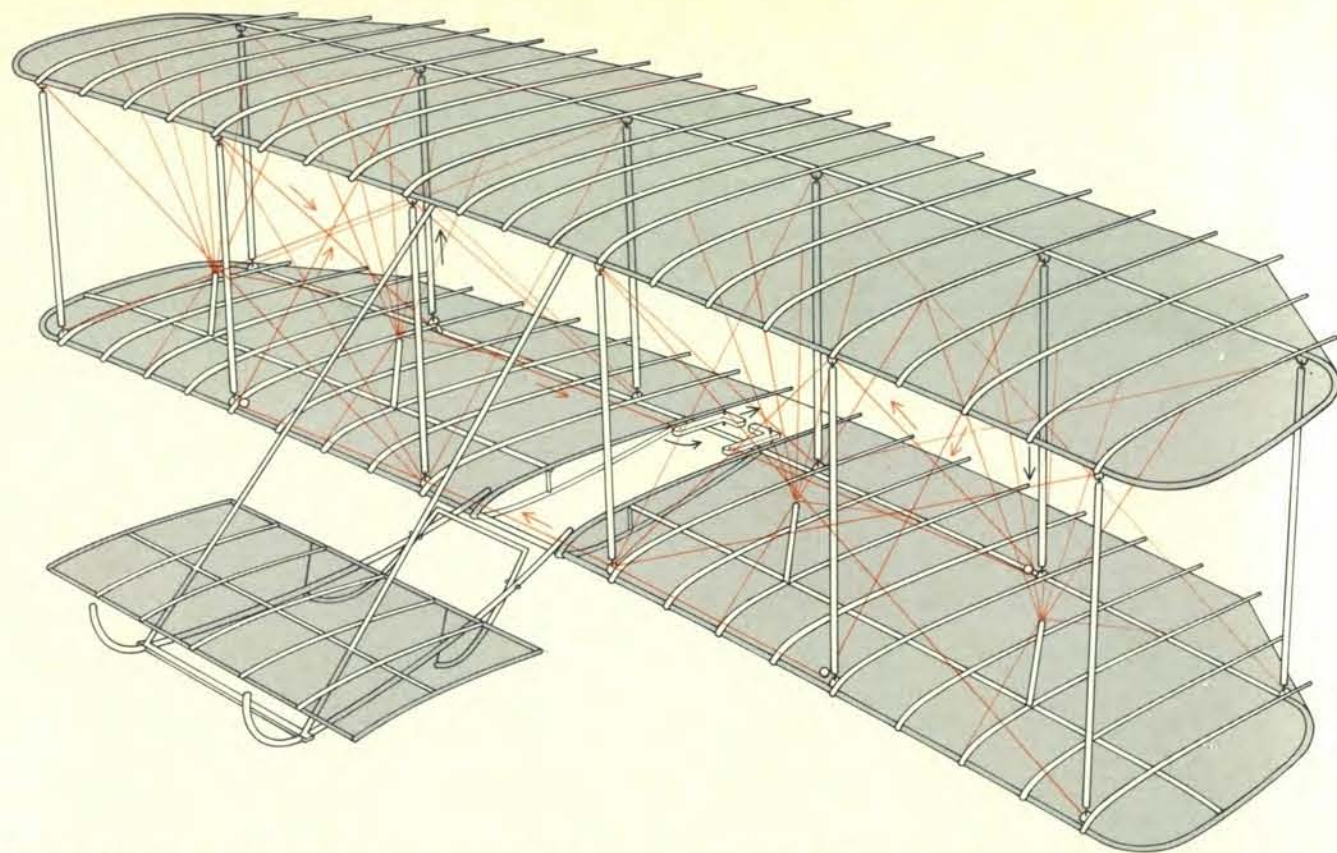
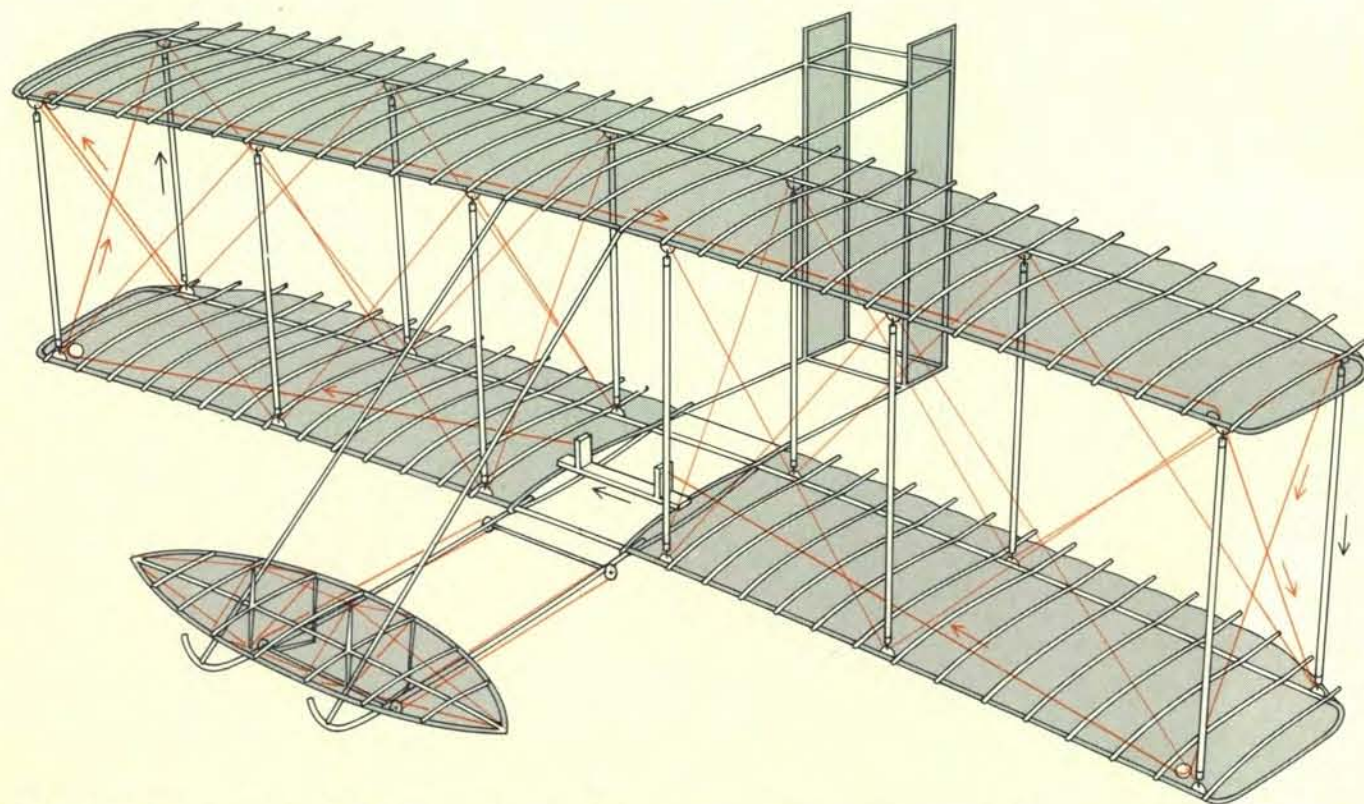
Ora il programma di prove si allarga. I fratelli costruirono un hangar di legno a Kill Devil Hills, sette chilometri da Kitty Hawk, dove le grandi dune di sabbia offrivano una buona piattaforma di partenza. A causa dei loro impegni di lavoro, condussero gli esperimenti di volo prima di quanto avevano programmato, dal 10 luglio al 20 agosto.

Risulta dai loro diari che Wilbur fece tutti i voli del 1901. Il primo giorno di prove, 17 luglio, eseguì 17 planate, senza aver prima provato il velivolo come aquilone. Questi primi voli furono pressocché disastrosi soprattutto per i problemi di controllo longitudinale. Wilbur scoprì che spesso era necessario deflettere completamente lo stabilizzatore orizzontale (il canard) e che la traiettoria di volo dell'aliante risultava notevolmente ondulata. Due volte l'angolo di incidenza fu così elevato che la macchina andò in stallo. Ogni volta, peraltro, l'aliante scivolò lentamente verso terra invece di picchiare bruscamente come era accaduto a Lilienthal. Queste due «riprese» del velivolo dalla condizione che Wilbur aveva temuto soddisfecero i fratelli e confermarono la validità della configurazione canard. E fino al 1910 essi non ebbero alcun interesse per l'impennaggio posteriore.

Wilbur aveva pur sempre il problema di capire perché l'aliante si comportasse così capricciosamente. Nessun miglioramento di rilievo fu ottenuto con la riduzione dell'area del piano di controllo da 18 a 10 piedi quadrati (meno di un metro quadrato). I fratelli fecero allora volare l'aliante come un aquilone e in un solo pomeriggio scoprirono che il centro di spinta si spostava all'indietro man mano che l'angolo di attacco diminuiva.

Si sapeva a quel tempo che per una superficie piana il centro di spinta si sposta continuamente dal centro, quando la superficie è perpendicolare al flusso di aria, verso il bordo di entrata (anteriore) quando l'angolo di incidenza è zero. Sperimentando un'ala i Wright scoprirono una proprietà fondamentale: la curvatura provoca lo spostamento del centro di spinta verso l'indietro se si cambia l'angolo di incidenza.

L'arretramento del centro di spinta fu la causa delle difficoltà che Wilbur ebbe nel controllo dell'aliante. La soluzione stava nella riduzione della curvatura delle ali. I fratelli tornarono alla curvatura dell'aliante del 1900: disposero dei supporti sull'ala inferiore con un'armatura di cavi per tirare la sezione centrale della centina verso il basso su tutte e due le ali. I risultati furono immediati e soddisfacenti. Wilbur fece una trentina di voli dopo le modificazioni: il più lungo durò 17 secondi e mezzo e la distanza massima fu di 117 metri.

Durante l'ultima settimana di prove

L'aliante del 1901 era il più grande mai sperimentato sino ad allora. Aveva un'apertura alare di 6,60 metri e una corda di 2,10 metri: il peso senza il pilota era di 45 chili. Nelle prove a Kitty Hawk, con Wilbur come pilota, l'aliante in principio segnò una linea di planata molto ondulata. I Wright modificarono l'aereo per ridurre la curvatura che avevano dato alle ali dal 1900: trovarono che così volava molto meglio.

L'aliante del 1902 consentì ai Wright di farsi la maggior parte della loro esperienza di volo. Fu il primo dei loro aerei ad avere il doppio timone verticale (aggiunto per compensare la tendenza della macchina a girarsi in una virata) e ad avere la slitta per il pilota. L'aereo aveva un'apertura alare di 9,60 metri e una corda di un metro e mezzo. I Wright avevano imparato dalle prove alla galleria del vento che un'ala lunga e sottile è più efficiente di una più corta e con la stessa superficie; inoltre, le ali avevano anche un diedro negativo, cioè erano piegate verso il basso.

Wilbur fece una fondamentale osservazione sul comportamento di un aereo quando viene fatto virare. Registrò nel suo diario che «l'ala che vira sembra cadere all'indietro, ma prima sale». Il 22 agosto scrisse a Chanute: «La settimana scorsa non portò grandi risultati benché noi si sia provato che la nostra macchina non vira (cioè gira) sempre verso l'ala più bassa, un risultato assai imprevisto e che rovescia completamente le nostre teorie circa le cause che portano alla virata a destra o a sinistra».

La virata di un aereo non si ottiene girando un timone ma rollando sull'asse longitudinale. Per virare a destra, per esempio, l'ala destra si abbassa e la sinistra si alza. La direzione della spinta totale (la portanza) è così deviata sulla destra della verticale e una parte della portanza contribuisce ad accelerare l'aereo verso destra. Furono i Wright a scoprire il modo corretto per virare, una scoperta resa possibile dalla loro invenzione dello svergolamento alare per mantenere la stabilità laterale.

Mentre Wilbur si esercitava alle virate scoprì quella che oggi è chiamata l'imbardata rovescia. Come inizialmente la portanza sull'ala alzata è maggiore che non su quella abbassata, così è della resistenza. Le diverse forze aerodinamiche tendono a far girare l'aeroplano in direzione opposta a quella che si voleva al momento di iniziare la rollata. Le osservazioni di Wilbur portarono alla installazione di un impennaggio verticale sul veleggiatore del 1902.

Poiché Wilbur era contemporaneamente un teorico, un inventore, un costruttore e un pilota collaudatore, in tre settimane di volo fu in grado di fare due decisive scoperte: la prima, l'influenza della curvatura alare sullo spostamento del centro di spinta, la seconda, l'imbardata rovescia. Le conseguenze che ne trasse completarono la configurazione dell'aereo.

La galleria del vento

Gli esperimenti del 1901 convinsero i Wright che i dati raccolti da Lilienthal sulle ali curve erano sbagliati. Misure più accurate confermarono ciò che avevano trovato nel 1900, e cioè che la previsione sul valore della portanza era in eccesso. A cominciare dall'autunno del 1901 i fratelli portarono avanti una lunga serie di prove in una galleria del vento.

La loro non era la prima galleria aerodinamica né nel mondo, né negli Stati Uniti. Probabilmente essi ricavarono l'idea da alcuni articoli di «Aeronautical Annual», una rivista pubblicata a Boston da James Means dal 1895 al 1897. Il contributo dei Wright come sempre riguardò l'essenziale. Essi raccolsero dati accurati per diverse forme alari o comunque superfici aerodinamiche. Non fecero in realtà misure assolute ma piuttosto compararono le forze che agiscono su un'ala con quelle che agiscono su una superficie piana perpendicolare al flusso d'aria. Come risultato delle loro prove, a metà inverno erano in possesso di tutti i dati necessari alla progettazione dei loro aerei per i dieci anni successivi.

I Wright decisero anche, sulla base delle proprie rilevazioni, che le misure fatte da Lilienthal non erano così accurate come avevano creduto in principio. Per le previsioni sulla portanza dei loro alianti a loro serviva una costante (riferita alla resistenza di una superficie piana) che era stata dedotta dai dati pubblicati da Smeaton. Paragonando i risultati ottenuti alla galleria del vento con le misure sulla portanza dei loro alianti, i Wright stabilirono che il valore della costante era sbagliato. L'errore aveva determinato una sopravvalutazione della portanza pari a circa il 40 per cento.

Il 28 agosto 1902, i Wright tornarono di nuovo a Kill Devil Hills con un altro aliante. Questo modello aveva un'apertura alare di 9 metri e 60 centimetri e una corda di un metro e mezzo, così che la superficie alare complessiva e il carico alare rimanevano sostanzialmente quelli dell'aliante del 1901. Dalle prove alla galleria aerodinamica i fratelli avevano imparato che un'ala lunga e stretta, con un più alto rapporto di allungamento, è più efficace di un'ala con la stessa superficie ma con un'apertura minore. La scelta del profilo alare fu pure decisa dalle prove alla galleria aerodinamica.

La macchina del 1902 fu la prima ad avere quella specie di culla che i fratelli progettarono per consentire al pilota in posizione prona di azionare i cavi dello svergolamento spostando lateralmente le anche. Anche il comando per muovere l'impennaggio orizzontale fu cambiato: fu installata una leva azionata con la mano sinistra.

Il cambiamento di configurazione più evidente e importante fu però un doppio impennaggio verticale, con una superficie totale di poco superiore a un metro quadrato. Questo impennaggio fu montato rigido sull'aliante del 1902 e non fu certo pensato per le variazioni di direzione. I Wright avevano capito bene che l'impennaggio verticale era necessario per contrastare l'imbardata rovescia.

Le prove dell'aliante in funzione di aquilone fecero subito capire ai fratelli che essi ora disponevano di un miglior aeroplano. Videro che con la nuova ala avevano molto migliorato il rapporto portanza-resistenza e che le difficoltà per il controllo longitudinale si erano ridotte. Il timone verticale fisso, poi, agiva in maniera da ridurre la tendenza all'imbardata in una virata.

Le prime prove con il velivolo utilizzato in funzione di aliante rivelarono una «risposta» eccessiva ai colpi di vento laterali, e perciò le ali furono montate con un diedro negativo. Le prove seguenti mostrarono che l'aliante del 1902 recava al pilota anche più noie, per il controllo laterale, che non la macchina del 1901. Quando l'aliante veniva scosso da un colpo di vento laterale, aveva una incontrollabile tendenza a oscillare in imbardata.

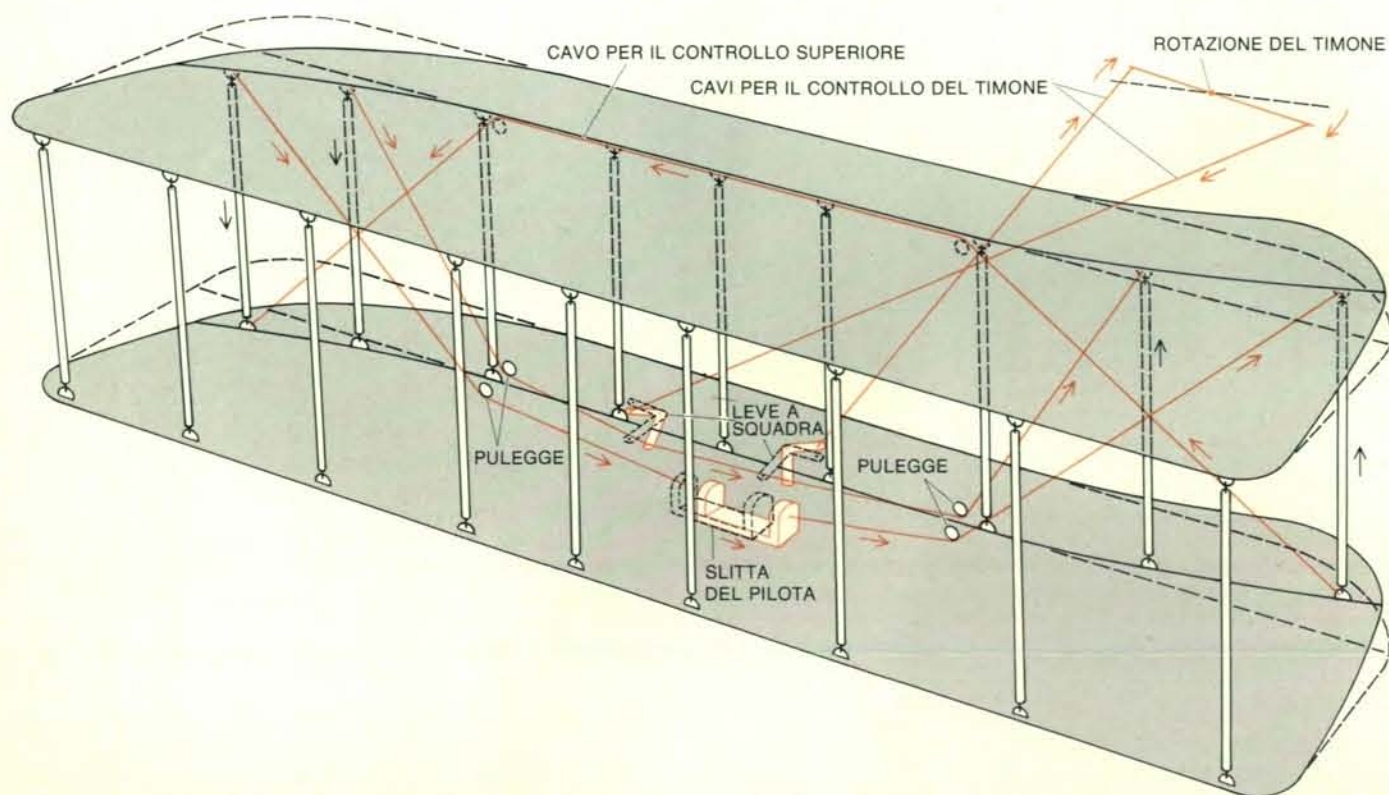
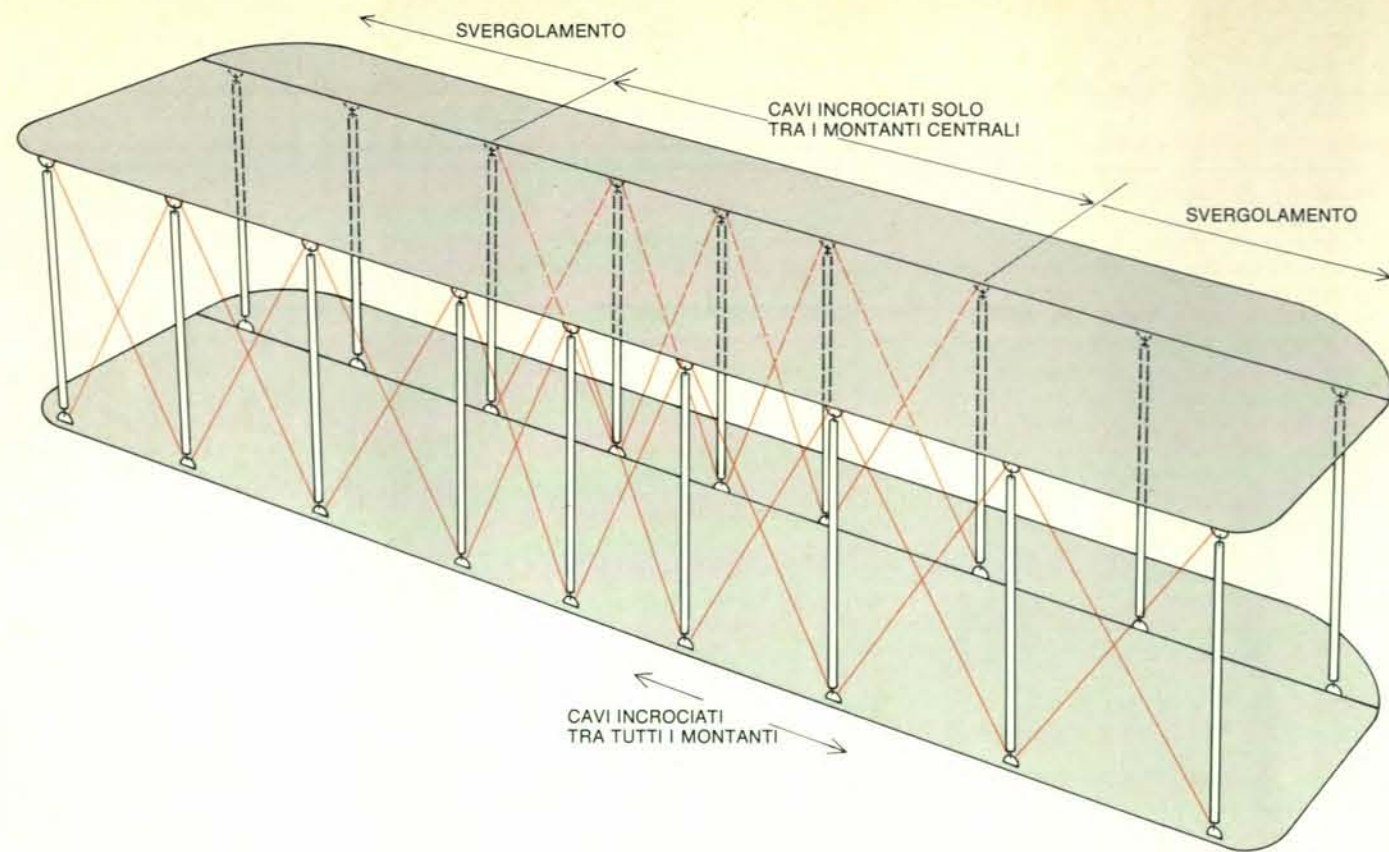
Problemi anche più seri si rivelarono in molte virate. Entrambi i fratelli sperimentarono l'inizio di una vite nel senso della virata. La sequenza stallo-vite è calcolata come la causa della perdita di controllo del 70 per cento degli incidenti aerei. I fratelli Wright furono i primi a scoprirlo. Essi incominciavano una virata, per esempio a destra, svergolando le estremità alari verso il basso a sinistra e verso l'alto a destra. La forza che agiva sul timone verticale tendeva essa pure a far girare l'aeroplano verso destra, contrastando l'imbardata rovescia. Se la virata diventava troppo «tagliata» (troppo a coltello), lo svergolamento veniva invertito. Questa manovra dava origine all'imbardata rovescia nella stessa direzione dell'effetto virata del timone verticale fisso. Per breve tempo l'aereo, quindi, virava anche più accentuamente verso destra: l'inizio di una vite. Poteva anche accadere che l'eccessivo svergolamento verso il basso delle ali all'interno della virata causasse il loro stallo con conseguente caduta. Diverse volte i fratelli dovettero constatare che questa serie di fatti terminava soltanto quando l'ala che era all'interno della virata faceva un solco nella sabbia, l'aeroplano si girava su se stesso e finalmente si fermava.

Fu un'idea di Orville di correggere la tendenza sostituendo il doppio timone fisso con uno singolo e mobile. Wilbur suggeriva che il suo funzionamento fosse collegato ai cavi dello svergolamento; quindi esso non avrebbe potuto essere mosso da solo. Tuttavia, una inversione dello svergolamento invertiva anche la tendenza all'imbardata prodotta dall'impennaggio mobile, e il risultato ottenuto fu una significativa riduzione nella tendenza alla vite.

Tutte queste diverse rettifiche completarono la configurazione e il sistema di comandi che i fratelli impiegarono sul Flyer del 1903. Non prima della fine del 1904 essi avrebbero «staccato» il timone verticale arrivando infine a ottenere un controllo indipendente del beccheggio, del rollio e dell'imbardata. Virtualmente tutti gli aeroplani moderni hanno lo stesso sistema di comandi.

Si prepara un motore

Le prospettive per la costruzione di una macchina volante apparivano assai più favorevoli dopo l'invenzione del motore a combustione interna sul finire del diciannovesimo secolo. Molta gente prima dei Wright aveva lavorato su questo problema, e il loro contemporaneo Samuel P. Langley, segretario della Smithsoniam Institution, arrivò, a 1903 avanzato, a sperimentare un aeroplano a motore che portava un uomo. In due prove molto pubblicizzate, il suo aereo lanciato con una catapulta da un battello sul Potomac non riuscì a volare e precipitò nel fiume. I progressi dei Wright non erano noti al pubblico, ma essi sentivano lo spirito della competizione.

Orville e Wilbur avevano incominciato a lavorare sul Flyer nell'autunno del 1902. Progettavano di installare un motore e delle eliche su una versione migliora-

Il sistema di svergolamento alare ideato dai Wright come mezzo per eseguire le virate. In alto, una vista frontale delle ali; ognuno degli otto spazi fra i montanti ha un incrocio di cavi, mentre soltanto quattro degli spazi posteriori hanno i cavi. A questo modo il bordo posteriore delle ali può essere mosso dai cavi agganciati alla «slitta» del pilota. Lo stesso sistema di cavi comanda anche le leve che muovono il timone posteriore verticale. Per eseguire una virata, per esempio a destra, il pilota doveva svergolare l'ala destra verso l'alto e l'ala sinistra verso il basso.

ta dell'aliante del 1902. Ritenevano che avrebbero potuto adattare un motore di automobile e che la letteratura tecnica marina avrebbe potuto fornire le basi per progettare eliche efficienti.

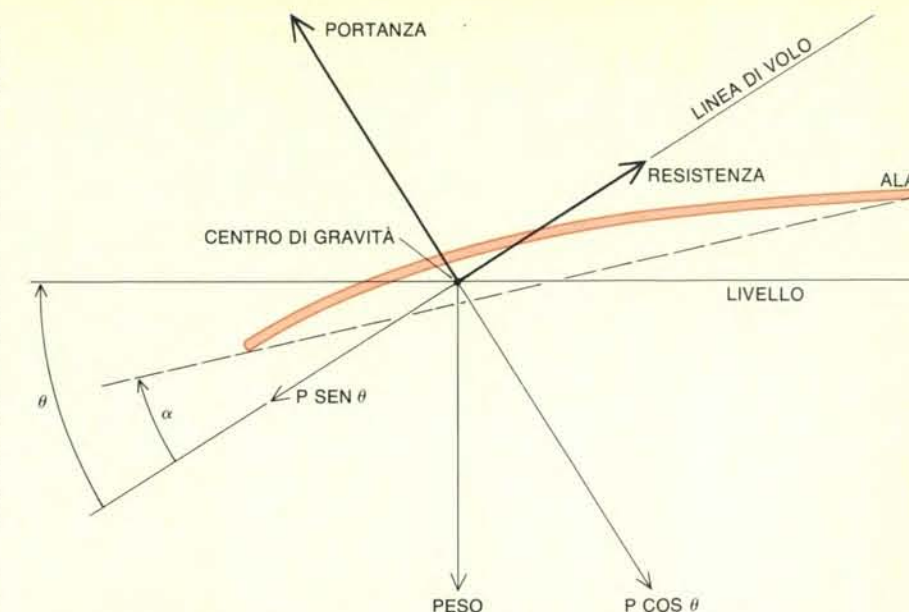
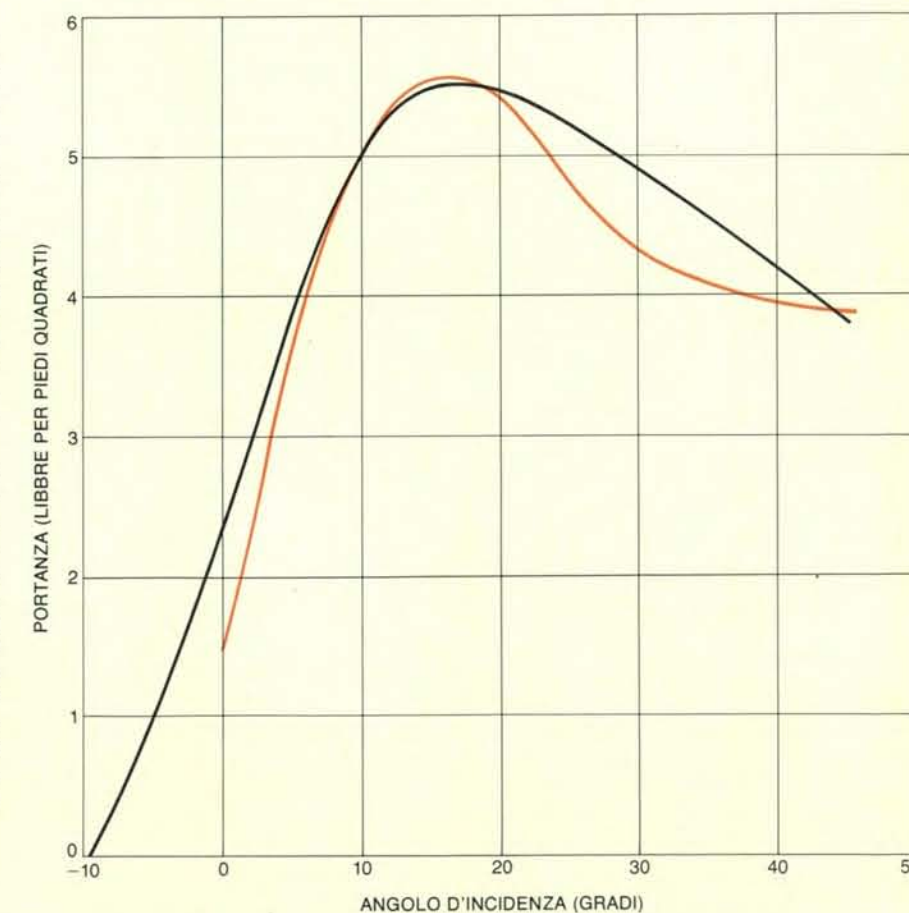
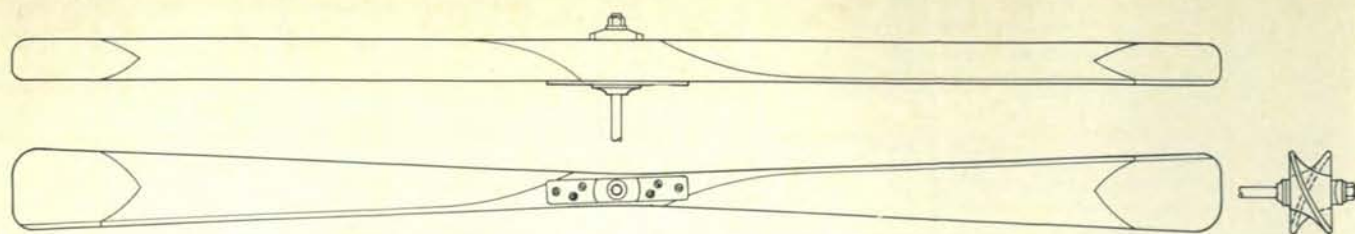
Dopo aver cercato invano di ottenere un motore dai costruttori di automobili decisero di progettarselo e di costruirselo da sé. Non fu una decisione particolarmente difficile per loro, perché avevano già costruito un motore a benzina per muovere le macchine nella loro fabbrica di biciclette. Dal marzo del 1903 disponevano di quello che serviva: un motore che sviluppava 12 cavalli-vapore e pesava 6,8 chilogrammi per cavallo-vapore. Non avevano cercato di creare un motore eccezionale, e il risultato fu effettivamente di gran lunga inferiore al motore costruito per Langley dal suo assistente ingegnere Charles Manly.

Subito dopo ebbero la sorpresa di scoprire che le teorie disponibili sulle eliche offrivano un solo dato utile: per una determinata potenza la migliore efficienza si ottiene facendo passare la maggior quantità possibile di aria attraverso l'elica (questo principio determinò il diametro piuttosto grande delle loro due eliche, 2 metri e 65). La conseguenza più notevole della loro scoperta fu che essi avrebbero dovuto elaborare un proprio metodo per progettare le eliche.

Non esistevano teorie per una progettazione non grossolana di qualunque tipo di elica. I Wright utilizzarono la loro conoscenza delle superfici alari per sviluppare quella che poi sarebbe stata conosciuta come teoria dell'elemento di pala, basata sull'idea che un piccolo segmento di una pala di elica va considerato come una sezione di ala. Poiché l'elica gira e l'aeroplano si muove, ognuna di queste piccole ali è esposta a velocità diverse e quindi genera diverse forze di portanza e di resistenza. La spinta dell'elica e la potenza da essa assorbita si trovano sommando le porzioni di portanza e di resistenza prodotte da ciascun segmento della pala.

L'analisi fatta dai Wright fece capire loro come progettare eliche con una efficienza del 70 per cento o anche maggiore in un'epoca in cui le eliche di altri ricercatori non avevano un'efficienza superiore al 50 per cento. Proprio la elevata efficienza delle loro eliche consentì ai fratelli di utilizzare, per il Flyer del 1903, il loro relativamente modesto motore. Il lavoro sulle eliche rappresentò il loro più grosso risultato in fatto di analisi, che però per essi valeva soltanto per le conseguenze pratiche. Essi non pubblicarono mai idee e analisi, e la moderna teoria delle eliche si rifà ad altre fonti.

Il primo volo

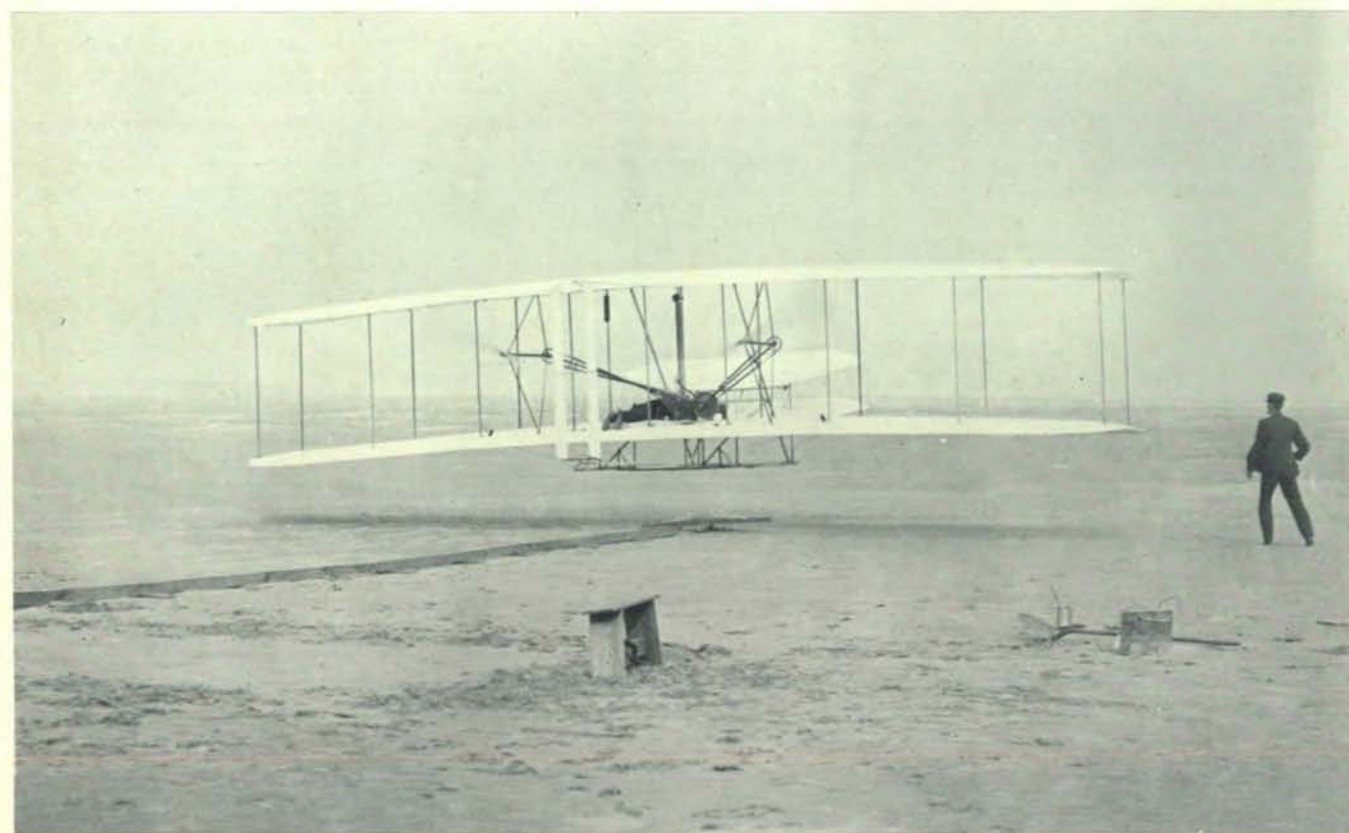
Il motore e le eliche furono montati su un velivolo che aveva la stessa configurazione dell'aliante del 1902, ma che era più grande, più robusto e perfezionato. Chiunque esamini l'aereo o i disegni di progetto resta impressionato dalla padronanza che i Wright avevano delle tec-

Le forze che agiscono su un aliante comprendono il peso (*P*), la portanza e la resistenza. Nel procedimento utilizzato intorno al 1900 per disegnare un'ala, si supponeva l'aliante in volo rettilineo (qui verso sinistra) a una velocità *V* lungo una traiettoria che fa un angolo θ sotto l'orizzonte. Una piccola parte del peso (P sen θ) tira l'aliante in avanti, equilibrando appena la resistenza. La portanza era definita come perpendicolare alla traiettoria di volo, «sostenendo» la componente del peso (P cos θ). Portanza e resistenza sono funzioni dell'angolo di incidenza (α).

Confronto fra dati sulla portanza ottenuti da Lilienthal *(in nero)* nel 1889 e dai Wright nel 1901 *(in colore)*. I Wright si trovarono di fronte al problema rivelato dalla loro prima fase intensiva di voli, con l'aliante del 1900; le loro ali generavano una portanza inferiore a quella prevista sulla scorta dei dati di Lilienthal. (Per la portanza sono stati mantenuti i valori originali, interessanti come documento storico.) Per correggere questo errore aumentarono la curvatura delle ali; poi fecero delle prove alla galleria del vento e i loro risultati si avvicinavano a quelli di Lilienthal quando l'angolo di incidenza dell'ala era tra cinque e dieci gradi, quindi la fonte dell'errore era altrove.

Il progetto di elica dei Wright, che essi realizzarono dopo aver scoperto che i dati disponibili sulle eliche marine offrivano scarso aiuto, si fondava sul principio che un piccolo segmento di pala poteva essere considerato come una sezione di ala. I dati navali suggerivano che la massima efficienza per una determinata quantità di potenza poteva essere ottenuta facendo passare la maggior quantità possibile di aria attraverso l'elica. Questo concetto condusse al notevole diametro (due metri e 65) delle due eliche alloggiate sul Flyer dei Wright (1903).

niche basate su legno, fili e tela. Il Flyer aveva un'apertura alare di 12 metri, una corda di 1 metro e 95 e una superficie totale di 47,381 metri quadrati. Il peso con pilota a bordo era di 339 chili e 750 grammi, il che dava un carico alare di 7,3 chilogrammi per metro quadrato.

Wilbur e Orville arrivarono a Kill Devil Hills il 25 settembre 1903, questa volta per starci parecchio, tre mesi. Ricostruirono l'hangar del 1902 e ne costruirono uno nuovo per il Flyer (13,20 metri per 4,80). Quando il tempo era buono affinavano il loro pilotaggio volando con l'aliante del 1902. Il resto del tempo lo passavano montando il nuovo aeroplano ed effettuando le prove preliminari su di esso.

In ottobre i fratelli seppero del primo insuccesso di Langley. La notizia poteva essere incoraggiante, ma anche le loro probabilità di successo erano marginali ed essi lo sapevano. Il loro apparecchio pesava undici chili più del previsto. L'aereo non era stato montato prima del viaggio a Kitty Hawks così che motore ed eliche non erano mai stati provati insieme alla cellula. Nella prima prova il motore girò con difficoltà, e poi si ruppero i supporti degli alberi delle eliche. Dovettero, forzatamente, mandarli a Dayton per ripararli, il 5 novembre.

Finché gli alberi delle eliche non furono rimandati indietro, il 20 novembre, i fratelli poterono fare ben poco. Il tempo scarseggiava, visto che Langley si preparava per un'altra prova come i Wright sapevano. Le condizioni meteorologiche stavano diventando un problema, la temperatura scendeva ogni tanto sotto zero di notte. Ma il 23 novembre erano state fatte prove sufficienti a misurare la spinta totale di cui potevano disporre: 60 chili. Poiché i fratelli avevano valutato la resistenza aerodinamica in 43 chili, ora confidavano che l'apparecchio potesse decollare e volare.

Mentre il motore girava per provare la strumentazione, il 28 novembre si ruppe l'albero di un'elica. Doveva essere sostituito e Orville ritornò nell'Ohio per costruire due nuovi alberi di materiale più resistente. Sul treno che da Dayton lo riportava a Kitty Hawk lesse su un giornale che la macchina volante di Langley era andata distrutta in un tentativo di decollo il giorno 8 dicembre. Non c'era più gara.

Il primo volo del Flyer del 1903 fu fotografato a Kill Devil Hills presso Kitty Hawk da John T. Daniels della Stazione di salvataggio di Kill Devil, che usò la macchina fotografica dei Wright. Il giorno era il 17 dicembre 1903. Orville è il pilota e Wilbur corre a lato. Il volo, uno dei quattro fatti dai Wright quel giorno, fu lungo 36 metri, la velocità media rispetto al suolo di 11 chilometri e mezzo all'ora. Di Wilbur l'ultimo e più lungo volo, 59 secondi. Flyer non fu più portato in aria e oggi è visibile all'Air and Space Museum della Smithsonian Institution.

La fauna degli argilloscisti di Burgess

I fossili di una formazione del Canada occidentale forniscono un ampio campionario d'una comunità animale vissuta nel Cambriano. Alcuni di questi animali sono antenati di forme attuali, altri sono unici e bizzarri

di Simon Conway Morris e H. B. Whittington

I fossili che rappresentano le prime, abbondanti testimonianze della vita sulla Terra sono di solito costituiti dalle parti dure di diversi animali marini invertebrati, ossia da conchiglie e da altri frammenti di scheletro esterno. Questo fenomeno tende a fornire un quadro generale dei fossili profondamente squilibrato. Basti pensare che dei circa 30 *phyla* di animali viventi attualmente, più della metà comprende specie dotate di poche parti dure o del tutto sprovviste di esse. Perciò l'evoluzione di questi phyla non è documentata quasi assolutamente da reperti fossili. Per fortuna la situazione attuale delle conoscenze non è del tutto squilibrata a favore degli animali forniti di scheletro. Sono stati scoperti alcuni depositi geologici che in seguito a circostanze eccezionali contengono fossili perfettamente conservati di animali dotati di corpo completamente o parzialmente molle. Qui descriveremo uno di questi depositi: gli argilloscisti di Burgess, del Canada occidentale. L'età antichissima e la ricca varietà di invertebrati marini negli argilloscisti di Burgess fanno sì che questo sia forse il più conosciuto di tutti i depositi di questo tipo. Oltre a descrivere la fauna di questa formazione, cercheremo di ricostruire il tipo di ambiente subacqueo in cui questi organismi vivevano, all'inizio dell'era paleozoica, circa 530 milioni d'anni fa.

Nell'autunno del 1909 il segretario della Smithsonian Institution, Charles Doolittle Walcott, stava compiendo ricerche sulle formazioni rocciose fossilifere della Columbia Britannica. Mentre seguiva un sentiero che attraversava il versante occidentale tra Wapta Mountain e Mount Field, nella parte meridionale della provincia, Walcott inciampò letteralmente sopra un blocco di argilloscisto caduto sul sentiero dal pendio sovrastante. Esaminando la roccia facilmente sfaldabile, egli fu meravigliato di trovare le impronte fossili di un certo numero di organismi dal corpo molle, conservati negli strati rocciosi. In una lettera a un collega di Toronto datata 27 novembre 1909, riferì in modo sintetico «di aver trascorso alcuni giorni per la raccolta... nelle vicinanze di Field e di aver trovato alcune cose interessantissime».

Walcott ritornò nello stesso luogo l'anno successivo per compiere ricerche sullo strato di argilloscisti che sovrastava il sentiero e che aveva lasciato cadere il blocco di roccia. La ricerca fu coronata da successo: trovò infatti due formazioni fossilifere separate tra loro da 21 metri di roccia di diverso tipo. In entrambe le formazioni compì uno scavo poco profondo e trovò che la più bassa era la più ricca di fossili. Inviò allora al distretto di Columbia migliaia di campioni fossili estratti da quello che egli stesso denominò «letto a fillopodi». (Il termine, oggi poco usato dai paleontologi, si riferisce a taluni artropodi fossili, probabilmente antenati degli attuali crostacei).

Come hanno dimostrato i lavori di Walcott e gli studi compiuti da altri paleontologi per decenni, i fossili degli argilloscisti di Burgess comprendono un numero assai rilevante di invertebrati marini: più di 120 specie. Alcuni di essi appartengono al phylum dei poriferi, ossia alle spugne. Questo phylum di animali primitivi è l'unico del sottoregno dei parazoi, un gruppo più evoluto del sottoregno dei protozoi, animali unicellulari, ma meno evoluto di quello dei metazoi, pluricellulari. Circa 10 altre specie rappresentano phyla di metazoi che erano sconosciuti prima della loro scoperta negli argilloscisti di Burgess; esse non sono presenti in altri reperti fossili. Le tracce di altre specie sprovviste di parti dure possono essere assegnate all'uno o all'altro phylum di metazoi i cui discendenti vivono ancora oggi e cioè:

Celenterati: phylum che comprende animali marini come meduse, pennatule e coralli. Le specie di celenterati degli argilloscisti di Burgess sono forse quattro.

Echinodermi: phylum comprendente, tra gli altri, stelle, ricci, gigli di mare e oloturie. Negli argilloscisti di Burgess sono state riconosciute almeno quattro specie di echinodermi.

Molluschi: phylum comprendente, tra gli altri, ostriche e vongole, calamari e polpi, e i primitivi chitoni (appartenenti alla classe degli anfineuri). Negli argilloscisti di Burgess sono state riconosciute tre specie di molluschi.

Artropodi: questo phylum comprende, tra le altre numerosissime specie, aragoste, gamberi, granchi e balani (tutti della classe dei crostacei) e, inoltre, un animale terrestre molto meno conosciuto, del genere *Peripatus* (membro della classe degli onicofori). Gli artropodi degli argilloscisti di Burgess comprendono parecchi rappresentanti della classe delle trilobiti, estinti da molto tempo, un animale di aspetto simile al *Peripatus* ma acquatico anziché terrestre, e circa 30 altre specie.

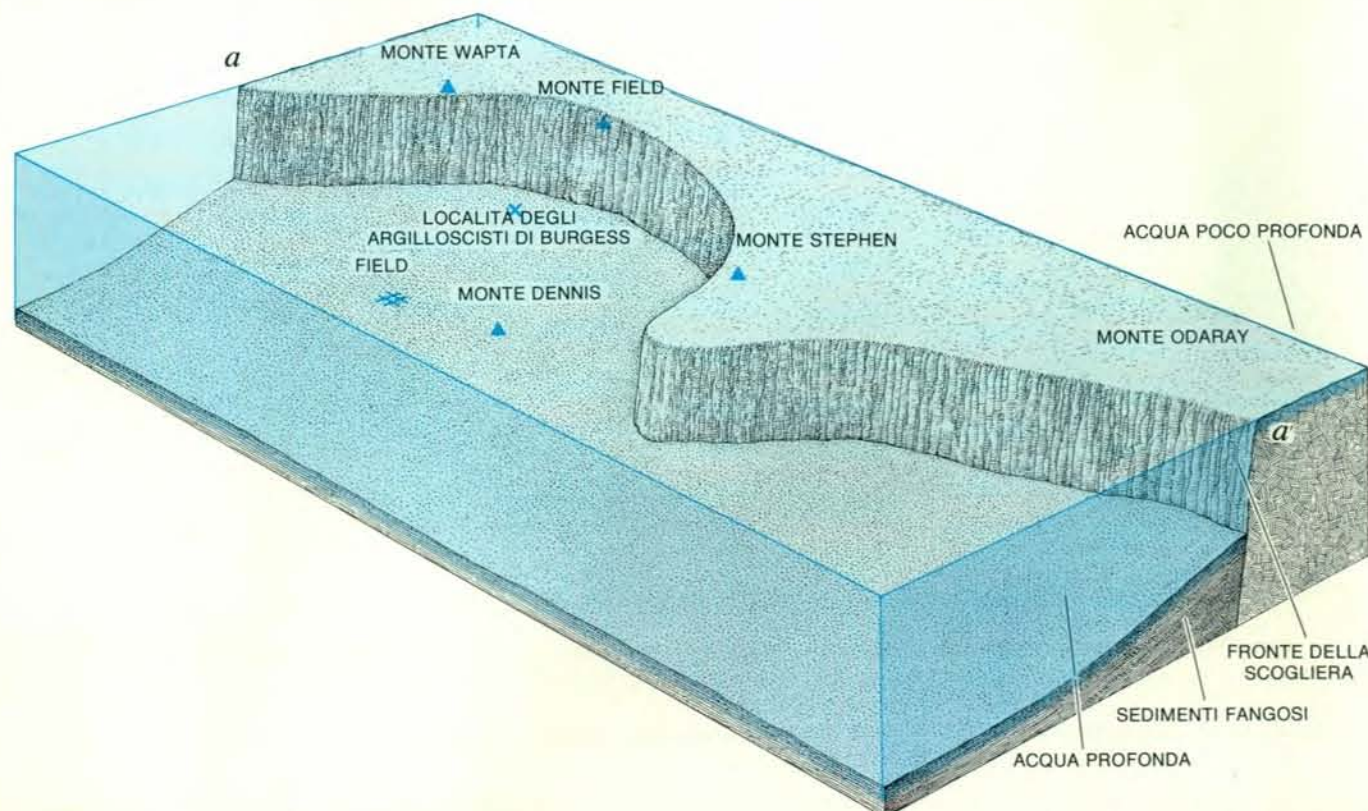
Priapulidi: un phylum minore di organismi vermiformi marini non metamerici. Dal genere vivente, *Priapulus*, deriva il nome del phylum. Sette specie di priapulidi, ora assai rari, prosperavano nei fondali marini fangosi che hanno dato origine agli argilloscisti di Burgess.

Anellidi: phylum che comprende lombrichi, sanguisughe e una classe di vermi marini, meno conosciuta ma molto vasta, i policheti. Il phylum degli anellidi è rappresentato, negli argilloscisti di Burgess, da sei specie.

Infine, nella fauna degli argilloscisti di Burgess si trova uno dei più antichi rappresentanti del phylum a cui anche l'uomo appartiene, i cordati. Tra i rappresentanti viventi di questo phylum, non appartenenti ai vertebrati, ricordiamo le ascidie e l'anfiosso. Negli argilloscisti di Burgess i cordati sono rappresentati dal genere *Pikaia*, con l'unica specie *P. gracilens*.

Questa fauna di animali dal corpo molle, così ben conservata, che rappresenta 8 phyla noti e 10 o più phyla prece-

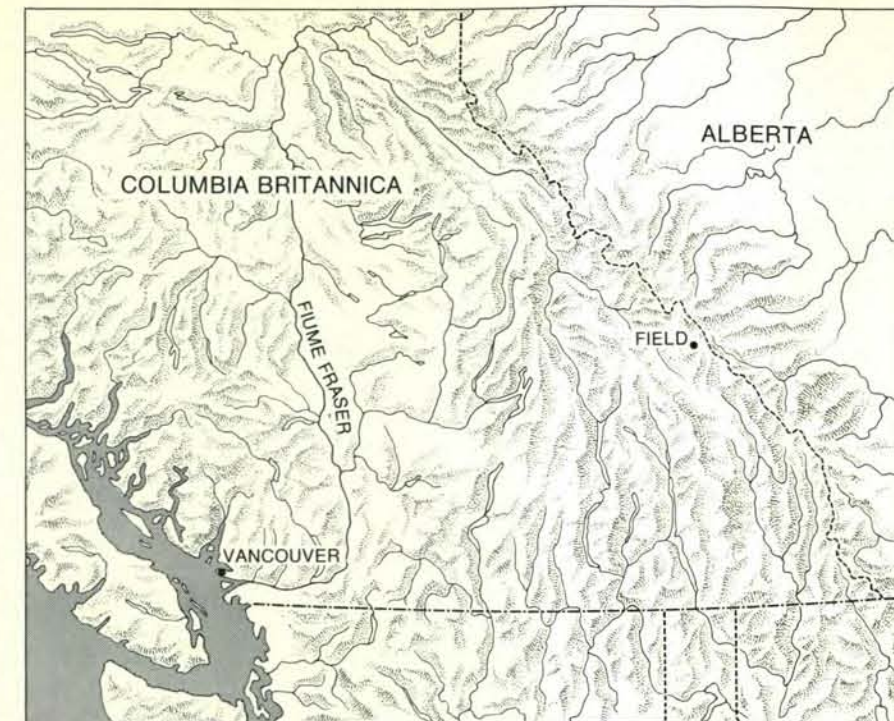
In queste fotografie di fossili si possono osservare quattro animali che vivevano nell'oceano nel Cambriano medio, circa 530 milioni d'anni fa. In alto a sinistra una trilobite, *Olenoides*, uno dei molti animali la cui anatomia è rimasta conservata nei più minuti particolari nel sedimento (silt) induritosi per formare gli argilloscisti di Burgess. L'esemplare è lungo 5,5 centimetri. A differenza di quasi tutti gli altri artropodi *Olenoides* possedeva arti non specializzati. In alto a destra un altro artropodo, *Waptia*. In posizione distesa questo animale, che si cibava sul fondale marino, era lungo 4 centimetri. In basso a sinistra vi è *Opabinia*, una delle circa 10 specie di animali trovati negli argilloscisti di Burgess che appartengono a phyla prima sconosciuti. Aveva cinque occhi, un corpo lungo 7 centimetri e nuotava vicino al fondale marino in cerca di cibo, grazie a una pinna caudale verticale. In basso a destra vi è uno dei molti organismi vermiformi non metamerici del fondale marino: *Selkirkia*, appartenente al phylum dei priapulidi attualmente rari. Con la sua proboscide sporgente misurava 5 centimetri.

La cava inferiore, denominata «letto a fillopodi» da Charles Doolittle Walcott che nel 1909 raccolse campioni di fossili negli argilloscisti di Burgess, mostra chiazze di neve. Questa cava e la zona più elevata di argilloscisti sono state riaperte nel periodo 1966-67 per cercare nuovi fossili.

L'affioramento di argilloscisti di Burgess, in questa ricostruzione del paesaggio marino del Cambriano medio, è una piccola porzione di un esteso deposito costituito da limo (più propriamente silt) accumulato ai piedi di una scogliera algale che formava una profonda insenatura e che s'innalzava verticalmente fino a un'altezza di circa 160 metri. La scogliera non sporgeva sopra il livello del mare, ma era ricoperta da uno strato d'acqua poco profonda. La scala verticale dello schema è esagerata di circa cinque volte, e la distanza (*a-a'*) da nord a sud lungo la scogliera è di circa 14 chilometri. Un sollevamento orogenetico e una dislocazione diedero origine alle Montagne Rocciose lungo il confine della provincia di Alberta e della Columbia Britannica, indicate coi triangoli colorati. La ricostruzione del fondale marino e della scogliera è basata su ricerche di I. A. McIlreath, uno degli studiosi che si sono occupati recentemente di questa singolare formazione.

dentemente sconosciuti e che prosperava nel Cambriano medio è già di per se stessa di grande interesse per i paleontologi. Tuttavia, oltre al loro interesse intrinseco, gli invertebrati degli argilloscisti di Burgess, con i loro adattamenti già specializzati, hanno un'importanza ancor più rilevante poiché gettano luce sulle prime fasi dell'evoluzione del regno animale. Gli unici animali a corpo molle di origine più antica sono rappresentanti del Precambriano e risalgono a 700-600 milioni d'anni fa, almeno 70 milioni d'anni prima del Cambriano medio: sono gli animali di Ediacara, scoperti circa 30 anni fa nelle colline di Ediacara nell'Australia meridionale e riconosciuti da allora in un certo numero di località sparse nel mondo. La fauna di Ediacara è in netto contrasto con la fauna degli argilloscisti di Burgess sia per i tipi di animali rappresentati (soprattutto celenterati), sia per la gamma piuttosto limitata di specializzazione di questi animali primitivi.

L'evento che separa la fauna di Ediacara dalla fauna degli argilloscisti di Burgess è una straordinaria diversificazione di animali pluricellulari che ebbe luogo pressapoco all'inizio del Cambriano. Perciò i fossili degli argilloscisti di Burgess ci permettono di constatare in un sol colpo i risultati di questo improvviso adattamento dei metazoi, in un periodo immediatamente successivo all'epoca in cui questo fatto si è verificato.

Nonostante il lavoro compiuto da Walcott e da altri, rimangono lacune notevoli nella conoscenza dell'ambiente in cui si sono formati gli argilloscisti di Burgess e del modo con cui si è conservata la fauna. Poiché ci si è potuti render conto appieno di queste lacune, si è capita l'importanza di una nuova ricerca compiuta nella località e più di dieci anni fa, a cura del Geological Survey del Canada, si sono ricominciati i lavori. Le autorità del Yoho National Park nella Columbia Britannica e l'amministrazione dei parchi canadesi a Ottawa hanno concesso un permesso speciale per raccogliere materiale dagli affioramenti fossiliferi. Gli scavi di Walcott furono riaperti nel 1966 e nel 1967 sotto la direzione di I. D. Aitken del Geological Survey del Canada. Quindi il nuovo materiale raccolto durante queste due stagioni, assieme a una parte della grande collezione sugli argilloscisti di Burgess ammassata da Walcott circa 60 anni prima, è stato inviato al nostro istituto presso l'Università di Cambridge per essere analizzato.

In quale ambiente viveva la fauna degli argilloscisti di Burgess? Grazie agli studi compiuti da I. A. McIlreath della Petro - Canada e da W. H. Fritz del Geological Survey del Canada, si è potuto dimostrare che gli animali vivevano sopra o entro un fondale melmoso, dove i sedimenti si erano accumulati alla base di una gigantesca scogliera. Questa struttura, costituita da materiale secreto dalle alghe, sorgeva verticalmente per qualche centinaio di metri da un bacino d'acqua profonda che veniva gradualmente riempiendosi di sedimenti. Gli affioramenti sparsi del fronte di tale scogliera si possono ancora rintracciare per chilometri e chilometri, attraverso la Columbia Britannica. Le acque del fondo del bacino avevano una circolazione evidentemente limitata ed erano ricche di idrogeno solforato e povere di ossigeno. I diversi invertebrati prosperavano dove i sedimenti melmosi erano accumulati contro la scogliera in modo da giungere a un'altezza superiore rispetto alle acque stagnanti del fondale, a circa 160 metri di profondità sotto il livello del mare.

I sedimenti situati davanti alla scogliera non erano stabili. Grazie a studi compiuti sugli argilloscisti da D. I. W. Piper della Dalhousie University, si è potuto dimostrare che avvenivano periodicamente crolli di materiale, per cui i sedimenti melmosi si spostavano nelle acque profonde e poco ossigenate del bacino. In questo modo venivano cancellate le impronte superficiali e le tane scavate sotto la superficie, formate dagli animali oggi inclusi negli argilloscisti di Burgess. Poiché gli animali travolti da questi torrenti di limo morivano quando restavano sepolti oppure appena dopo, non potevano lasciare nuove tracce. Ciò significa che il modo di vita di ciascuna specie deve essere dedotto da uno studio degli organi di locomozione e da confronti con gli invertebrati del medesimo tipo viventi attualmente.

Nello stesso tempo queste frane, seppellendo gli animali in depositi anaerobici di fine «silt», dove neppure gli animali che si cibano di organismi in decomposizione possono sopravvivere, favorivano enormemente la conservazione delle parti molli degli organismi. A mano a mano che il limo si induriva e si trasformava in roccia compatta, le carcasse sepolte risultavano appiattite e le parti molli venivano trasformate in sottili pellicole di allumosilicato di calcio. In genere le pellicole sono piuttosto scure, ma alcune parti di molti esemplari sono visibili come superfici altamente riflettenti.

Paradossalmente, benché le parti molli degli animali si siano conservate in modo quasi perfetto, spesso si possono scoprire segni di putrefazione dopo il seppellimento. Molti esemplari sono associati a un'area macchiata di nero, prodotta dal fatto che il contenuto corporeo della carcassa si è diffuso nel fango circostante. Nei casi estremi il fossile d'un verme è costituito solo da un sacchetto cavo, fatto di cuticola, poiché praticamente tutti gli organi interni dell'animale si sono dissolti per decomposizione. In certi vermi un segno quasi impercettibile della decomposizione è fornito dal distacco dei muscoli della parete corporea dalla cuticola.

Il seppellimento in una corrente melmosa ha altre conseguenze importanti. Da un lato, molti animali si trovano sepolti secondo diverse angolature; il letto di argilloscisto li ha perciò conservati in una grande varietà d'orientamenti, che rivelano un numero maggiore di strutture anatomiche rispetto a una semplice sepoltura orizzontale. D'altra parte, i sedimenti fluidi che penetrano tra le appendici di animali come gli artropodi e i policheti trasportati da una corrente turbolenta di silt si riducono alla fine in sottili strati d'argilloscisto. Se il ricercatore riesce a rimuovere questi straterelli lavorando pazientemente con un microcesello, può

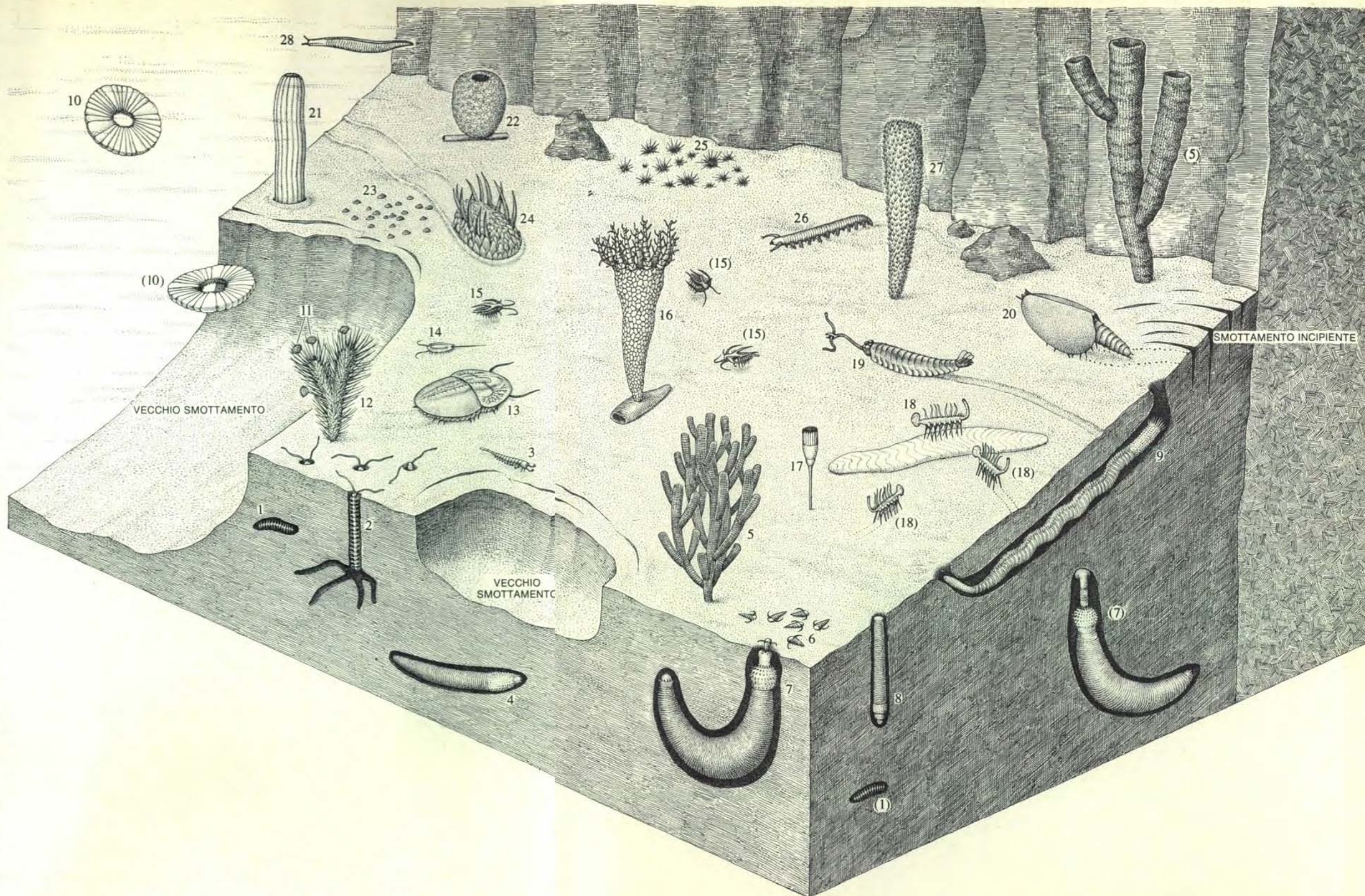
La formazione di argilloscisti di Burgess è situata a 560 chilometri a nordest di Vancouver, presso Field, nella Columbia Britannica. La formazione fossilifera fu scoperta da Walcott nel 1909.

rivelare particolari ulteriori dell'anatomia del campione, che altrimenti rimarrebbero nascosti.

La composizione della fauna degli argilloscisti di Burgess rivoluziona le nozioni convenzionali su quello che costituisce un tipico raggruppamento di animali del Cambriano. I fossili scoperti nella maggior parte delle località cambriane sono esoscheletri di artropodi del tipo delle trilobiti, conchiglie di vari membri del phylum dei brachiopodi e resti di echinodermi, per esempio appartenenti alla classe estinta degli eucrinoidi, dotati di guscio a placche. Invece negli argilloscisti di Burgess gli animali di questo tipo costituiscono appena il 20 per cento dei generi di invertebrati. Possiamo perciò considerare l'insieme di animali di questa formazione come un gruppo normale per il periodo cambriano, almeno per quanto riguarda la fauna delle acque profonde, e considerare le altre comunità animali del Cambriano squilibrate nelle percentuali a causa della fossilizzazione selettiva degli animali dotati di parti dure?

Poiché gli argilloscisti di Burgess rappresentano un unico ambiente congelato per un attimo del tempo geologico, non si può dare una risposta sicura alla domanda precedente. Tuttavia esistono parecchie prove secondo cui la fauna di tale formazione non sarebbe insolita per il Cambriano medio. I ritrovamenti di specie simili a quelle degli argilloscisti di Burgess in altre rocce cambriane sono a favore dell'esistenza in questo periodo d'una fauna assai diffusa di animali a corpo molle. Inoltre, in alcuni letti fossiliferi del Cambriano, determinate specie piuttosto singolari poterono prosperare poiché l'accesso del mare all'area di deposito era limitato. Gli argilloscisti di Burgess, d'altra parte, giacciono sulla sponda del mare aperto e sarebbero stati esposti alla colonizzazione di larve marine portate dalle correnti in quella località da altre zone. Questa circostanza aggiunge verosimiglianza all'ipotesi che la fauna degli argilloscisti di Burgess sia abbastanza rappresentativa di quella del periodo Cambriano.

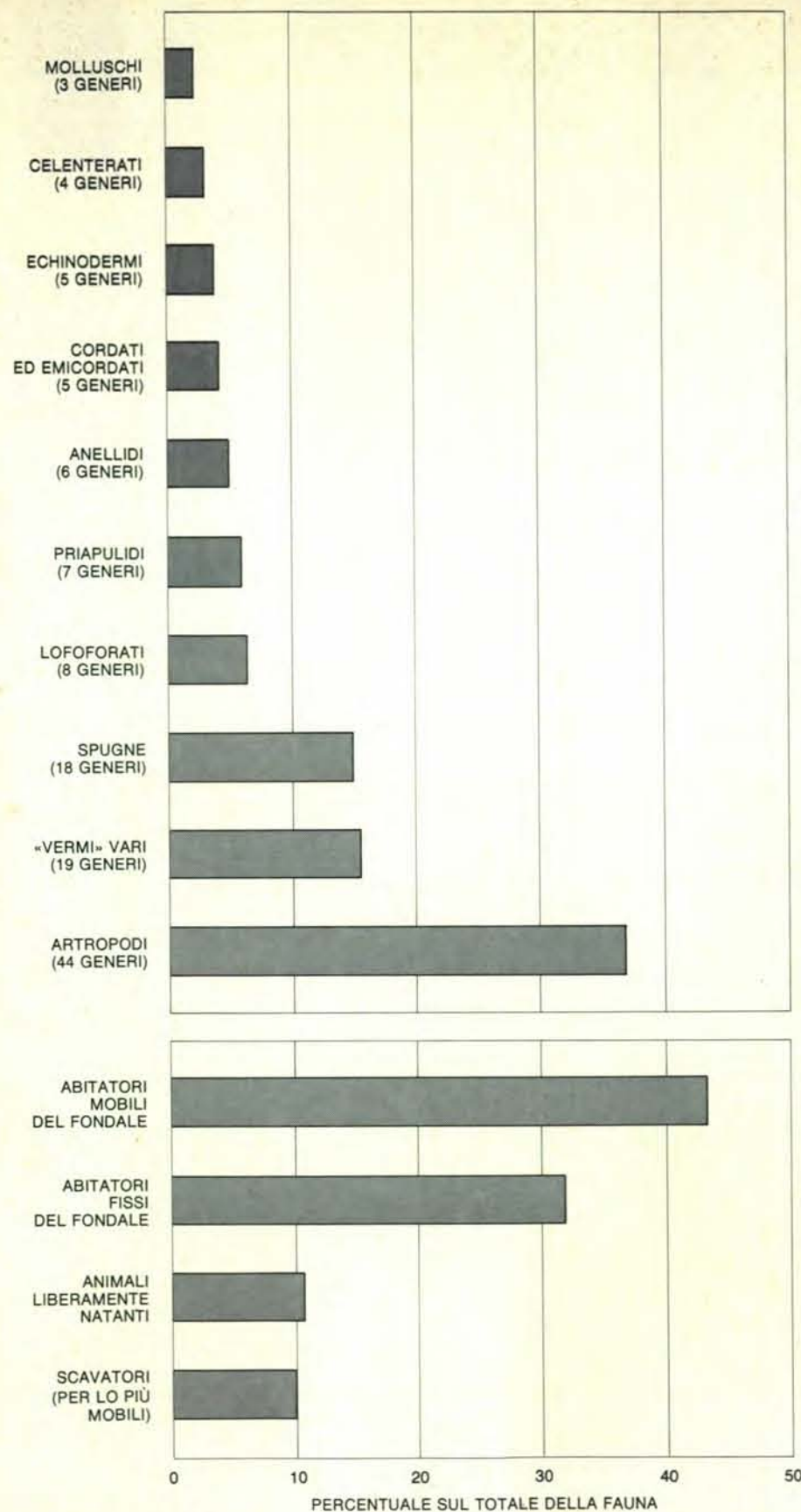
Sotto questo aspetto si dovrebbe notare che i rappresentanti di alcuni invertebrati attuali, che hanno avuto quasi certamente una lunga storia geologica, sono assenti dagli argilloscisti di Burgess. Non è presente alcuna specie del phylum dei platelminti, ossia dei vermi piatti, che comprendono fasciole, tenie e planarie. Non esistono neppure specie del phylum dei nemertini e del phylum dei sipunculidi. Forse tali organismi marini non sono rappresentati semplicemente perché l'ambiente di fronte alla scogliera non era adatto a essi.

Le specie che si trovano negli argilloscisti di Burgess possono essere classificate come membri delle comunità ecologiche marine viventi sul fondale, che prosperavano sul fango di fondo negli intervalli compresi tra due frane successive. Il fango ospitava un gruppo molto attivo di

In questa ipotetica ricostruzione sottomarina è raffigurato il luogo in cui i sedimenti siltosi formavano un pendio in discesa a partire dai piedi della grande scogliera, ambiente nel quale viveva la fauna degli argilloscisti di Burgess. Gli animali non sono raffigurati nelle medesime proporzioni numeriche dei reperti fossili e sono identificati da un numero; nell'illustrazione è raffigurato solamente circa un quinto delle specie fossilizzate. La maggior parte degli animali fissi sul fondale marino sono spugne: *Pirania (12)*, con brachiopodi attaccati alle spicole; *Eiffelia (22)*; *Choia (25)*, animale gregario; una specie gracile di *Vauxia (5)*, con una specie più robusta in alto a destra, e *Chancelloria (27)*. Tre altri animali immobili sono *Mackenzia (21)*, un celenterato; *Echmatocrinus (16)*, un crinoide primitivo, attaccato a un tubo vuoto di un verme, e *Dinomischus (17)*, una delle specie che rappresentano phyla di invertebrati finora sconosciuti. Gli animali che vivevano nascosti in tane sono *Peronochaeta (1)*, un polichete che si cibava di particelle di cibo nel limo; *Burgessochaeta (2)*, un secondo polichete che catturava il cibo coi lunghi tentacoli; *Ancalagon (4)*, un priapulide forse antenato di alcuni parassiti attuali; *Ottoia (7)*, un altro priapulide, che si sta cibando del mollusco *Hyolithes (6)*, mentre a destra sta scavando una tana; *Selkirkia (8)*, un terzo priapulide, in una tana, con l'estremità anteriore verso il basso, e *Louisella (9)*, un quarto priapulide che viveva in una tana a doppia entrata e muoveva il corpo per sospingere l'acqua ossigenata sopra le branchie. *Peytoia (10)* è un celenterato natante a forma di fetta d'ananas. Oltre a *Hyolithes*, altri molluschi del fondale marino sono *Scenella (23)*, le cui parti molli sono coperte da una conchiglia a forma di cappello cinese, e *Wiwaxia (24)*, coperta di squame e spine difensive. Tra i numerosi artropodi del fondale marino sono raffigurati: *Yohoia (3)*, con le sue caratteristiche appendici prensili; *Naraoia (13)*, una trilobite atipica che manteneva caratteristiche larvali; *Burgessia (14)*, con la sua lunga spina caudale; *Marrella (15)*, che forse nuotava appena sopra il fondale marino; *Canadaspis (20)*, un crostaceo primitivo, e *Aysheaia (26)*, un animale dalle zampe tozze che ricorda *Peripatus*, un abitatore attuale della terraferma. Altri rappresentanti di nuovi phyla oltre a *Dinomischus* sono *Hallucigenia (18)*, che si prepara a cibarsi di un verme morto, e *Opabinia (19)*, nell'atto di afferrare un vermetto con la sua unica appendice biforcuta. Infine si può osservare *Pikaia (28)*, unico rappresentante del phylum dei cordati nel Cambriano medio, mentre nuota solitario. Probabilmente usava la sua serie di muscoli a zig-zag per spostarsi al di sopra del fondale marino. Il phylum dei cordati comprende il subphylum dei vertebrati, evolutosi più tardi.

invertebrati scavatori, tra cui predominavano i priapulidi. Attaccate al fondale e sviluppate fino a diverse altezze vi erano diverse spugne, rappresentate da almeno 15 generi; queste si nutrivano di particelle di cibo sospese in acqua. Molte specie di artropodi perlustravano attivamente la superficie del fondale marino o scavavano nel fango in cerca di cibo. Alcuni brachiopodi occupavano una nicchia particolare: si attaccavano alle spicole allungate di una delle specie di spugne, *Pirania*. Per i brachiopodi i vantaggi sono evidenti: vivevano infatti al di sopra delle acque torbide del fondale e, a questi livelli più elevati, potevano catturare particelle di cibo come quelle di cui si cibavano le spugne.

Oltre a questa comunità di abitatori di superficie, fissi o mobili, e di scavatori, un certo numero di specie liberamente natanti abitava le acque davanti alla parete formata dalla scogliera. Poiché alcuni rari esemplari rimanevano sepolti accidentalmente nei sedimenti franati, su questa comunità animale si può solo gettare un rapido sguardo, anche se molto istruttivo. I diversi membri di questa fauna pelagica probabilmente vivevano a diverse profondità; alcuni tra essi possono essere stati gettati dal mare aperto nella zona della scogliera.

Di solito nelle località fossilifere del Cambriano la maggior parte dei fossili è costituita dagli esoscheletri mineralizzati delle trilobiti, gli artropodi paleozoici più conosciuti. Invece negli argilloscisti di Burgess le trilobiti sono relativamente poco rappresentate, con un'eccezione: *Olenoides*, che assume un significato assai importante perché in parecchi esemplari le appendici sono state ben conservate e appaiono con ricchezza di particolari. *Olenoides* era fornito di un paio di sottili antenne nella parte anteriore e un paio di cerci, strutture simili ad antenne, nella parte posteriore del corpo. Gli arti disposti lungo i lati dell'animale, fino a 16, erano tutti simili per struttura. La coxa, il segmento più sviluppato, prossimale al corpo, portava una batteria di temibili spine. Attaccate alla coxa vi erano due appendici: una era una branchia filamentosa, l'altra una zampa, o appendice locomotoria. *Olenoides* poteva afferrare e fare a pezzi animaletti molli, come piccoli vermi, poi ne portava alla bocca i frammenti. Le antenne anteriori e i cerci posteriori indubbiamente servivano all'animale per raccogliere informazioni sulla presenza di cibo e di predatori potenziali.

I quattro animali raffigurati nelle fotografie a sinistra sono i più singolari della fauna di Burgess. In alto si vede *Hallucigenia*, già rappresentata nell'illustrazione delle due pagine precedenti. In seconda posizione dall'alto vi è *Nectocaris*, un animale affusolato dotato di notevoli raggi della pinna. In basso a sinistra è raffigurata *Amiskwia*, un verme gelatinoso con pinne sporgenti; in basso a destra *Dinomischus*, animale peduncolato la cui forma è ricostruita nell'illustrazione delle due pagine precedenti. Queste quattro specie e sei o più altre rappresentano phyla d'invertebrati finora sconosciuti.

I generi degli argilloscisti di Burgess finora riconosciuti sono 119. Nella parte superiore di questo diagramma viene indicata la percentuale del totale assegnato ai vari phyla. Circa il 40 per cento del totale è costituito da artropodi; solo 14 su 44 generi d'artropodi sono trilobiti. Gli organismi vermiformi non appartenenti ai priapulidi e agli anellidi (19 generi), assieme alle spugne (18 generi), costituiscono il 30 per cento del totale; i molluschi sono quelli più scarsamente rappresentati. In termini di habitat, come indica la serie inferiore di istogrammi, più del 40 per cento degli animali degli argilloscisti di Burgess si spostava sul fondale marino e più del 30 per cento era fissile. La maggior parte degli animali rintanati si spostava liberamente, mentre alcuni rimanevano immobili.

Il fatto che gli arti primitivi di questa trilobite siano tutti simili è in netto contrasto con l'organizzazione del corpo di molti artropodi fossili e viventi, i cui arti sono variamente modificati e specializzati.

Circa il 40 per cento della fauna degli argilloscisti di Burgess è costituito da artropodi. Sia per il numero di specie, sia per il numero di esemplari singoli, i rappresentanti dal corpo molle di questo phylum superano le trilobiti, dotate di guscio duro. Molti di questi «non-trilobiti» hanno appendici conservate nei più minuti particolari; alcuni di essi devono essere stati predatori o necrofagi efficientissimi. Il genere più abbondante è *Marrella,* un artropodo dalla testa a forma di cuneo, che portava due paia di lunghe spine a forma di corno, con la curva rivolta all'indietro. *Marrella* esplorava l'ambiente del fondale marino con un paio di antenne e raccoglieva particelle di cibo portandole alla bocca con le appendici piumose adiacenti. Le 20 o più appendici laterali erano articolate: ciascuna dotata di una ramificazione costituita da una branchia filamentosa.

Il secondo genere in ordine di numero fra gli artropodi non-trilobiti è *Canadaspis.* L'animale era nascosto quasi del tutto da una corazza divisa in due pezzi, che lasciava scoperta solo la parte anteriore del corpo. Asportando con attenzione questa corazza, appaiono le appendici, notevolmente simili a quelle di alcuni crostacei viventi. Anche uno dei più grossi predatori del fondale marino era un artropodo: ci riferiamo a *Sidneya,* un animale le cui caratteristiche appendici ricordano quelle del limulo, un grosso artropodo tuttora vivente. È stato possibile identificare parti del contenuto intestinale di *Sidneya,* costituite da frammenti di conchiglie di brachiopodi, il che sta a dimostrare come questo animale fosse in grado di spezzare il duro guscio delle sue prede.

Sidneya non era il più grosso tra gli artropodi degli argilloscisti di Burgess. La roccia contiene le impronte di grandi zampe isolate; se sono proporzionate al resto del corpo, queste zampe devono essere appartenute a un animale lungo circa un metro. Questi resti sono stati attribuiti al genere *Anomalocaris,* che probabilmente rappresenta uno dei più grandi invertebrati del periodo Cambriano.

Il più interessante artropodo di Burgess è *Aysheaia,* un animale dal corpo tarchiato e dalle zampe tozze. Quando Walcott pubblicò per la prima volta una fotografia di questo fossile mezzo secolo fa, un certo numero di zoologi gli scrisse per far notare quanto questo invertebrato del Cambriano medio assomigliasse a *Peripatus,* un animale con otto generi affini (comprendenti due famiglie) della piccola classe degli onicofori, appartenente al phylum degli artropodi. *Peripatus* è un animale terrestre, mentre *Aysheaia* era una forma marina; tuttavia, *Aysheaia* certamente rappresenta il tipo di animale ancestrale da cui potrebbero essersi evoluti alcuni artropodi viventi, come i miriapodi e gli insetti.

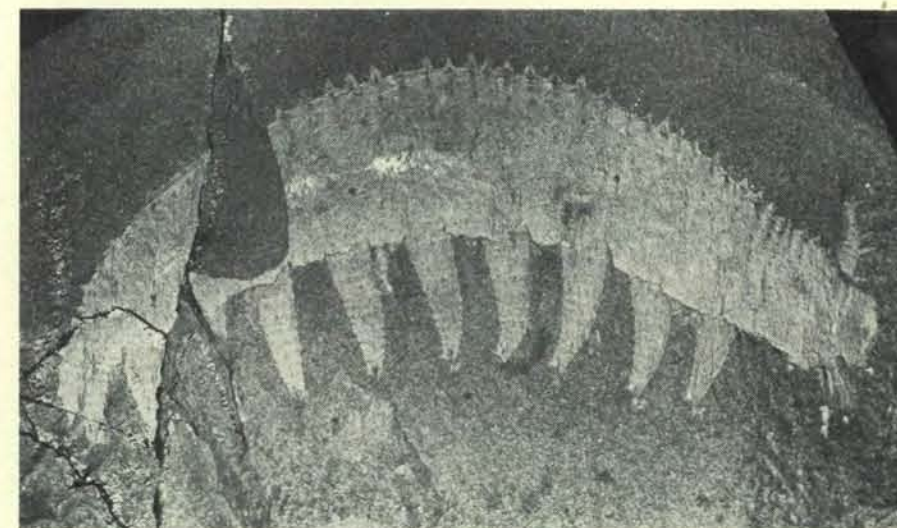
Benché alcuni degli artropodi non-trilobiti negli argilloscisti di Burgess, come *Aysheaia,* ricordino forme che sarebbero apparse più tardi, la maggior parte di essi non può essere situata in alcun gruppo riconosciuto. Questi animali non hanno affinità chiara né con le altre specie degli argilloscisti di Burgess, né con gli artropodi dei periodi successivi. Poiché possedevano un numero notevolmente alto di caratteristiche anatomiche, che denotano un alto grado di specializzazione, possiamo dedurre che nel periodo cambriano gli artropodi erano in una fase di radiazione adattativa di cui finora non si sospettava l'esistenza. I numerosi ceppi sorti durante questo periodo di rapida evoluzione non sembrano aver lasciato per lo più discendenza alcuna. È interessante notare che gli animali in procinto di diventare dominanti nei successivi periodi geologici in genere hanno una posizione abbastanza insignificante nelle varie faune cambriane; un osservatore ipotetico si sarebbe trovato in difficoltà se avesse voluto fare una previsione a lunga scadenza sul successo biologico dei vari gruppi d'animali.

Due phyla di organismi vermiformi della fauna degli argilloscisti di Burgess sono costituiti dai notissimi anellidi e dai meno noti priapulidi. Un tipico anellide della formazione è *Canadia,* un polichete *(in alto):* le sue *setae,* organi di locomozione, sono conservate nei particolari. Un tipico priapulide *(in basso)* è *Louisella* che appare nella ricostruzione delle pagine 76 e 77.

Di tutti gli animali degli argilloscisti di Burgess, a eccezione degli artropodi, sono particolarmente degni di nota i rappresentanti di sei phyla. Tra gli echinodermi una volta si pensava che fosse ampiamente rappresentata la classe delle oloturie, volgarmente dette cetrioli di mare. Ora solo uno tra questi animali, e precisamente *Eldonia,* ha tale collocazione sistematica. A differenza della stragrande maggioranza delle specie appartenenti alla classe, *Eldonia* aveva un corpo medusoide e un paio di tentacoli orali. Questi animali probabilmente nuotavano attraverso l'acqua in branchi, usando i tentacoli per catturare il cibo. Un altro echinoderma degli argilloscisti di Burgess, il giglio di mare *Echmatocrinus,* è il più antico crinoide che abbia lasciato tracce fossili e ovviamente mostra un certo numero di caratteri primitivi.

Le specie appartenenti al phylum dei celenterati sono tra le più primitive dei metazoi. Per esempio, nel gruppo di fossili di Ediacara, del tardo Precambriano, i celenterati predominano. Invece i diversi celenterati degli argilloscisti di Burgess, alcuni simili alle meduse, altri alle pennatule, sembrano aver avuto un ruolo piuttosto limitato nella comunità. D'altro canto le spugne, gli animali più primitivi della medesima formazione, erano membri importanti della comunità: erano infatti numerose e di forma svariata; alcune specie crescevano fitte sul fondale al punto da formare veri ammassi.

I vari «vermi» degli argilloscisti di Burgess furono assegnati da Walcott principalmente al phylum degli anellidi in generale e alla classe dei policheti in particolare. Ora si sa con certezza che molti di questi animali appartengono ad altri phyla. Tuttavia è proprio tra i policheti che si possono trovare alcuni tra gli esempi più spettacolari di conservazione di corpi molli: le *setae,* ciuffi di fini setole, usate come organo di locomozione da parte di questi animali, sono rimaste conservate particolarmente bene, come pellicole brillanti e riflettenti negli argilloscisti. Uno dei policheti, *Canadia,* non viveva evidentemente in una tana, ma per lo più nuotava in prossimità del fondale marino. Un altro, *Burgessochaeta,* era molto probabilmente un tipico scavatore il quale trovava rifugio nel fondale fangoso e andava in cerca di cibo attorno all'entrata della tana servendosi dei lunghi tentacoli.

Oggi il phylum dei priapulidi interessa solo un ristretto gruppo di specialisti. Durante il periodo cambriano invece questi vermi costituivano un gruppo importante, e due priapulidi presenti negli argilloscisti di Burgess sono particolarmente degni di menzione. Uno, *Ottoia,* è il più numeroso del gruppo. È in ottimo stato di conservazione, fin nei minimi par-

Questo invertebrato, *Aysheaia,* notevolmente ben conservato, è un artropodo ancestrale dotato di una straordinaria rassomiglianza con *Peripatus,* un onicoforo attualmente vivente. Alcuni artropodi degli argilloscisti di Burgess del tipo di questa *Aysheaia* sono stati probabilmente gli antenati dei miriapodi e degli insetti, rappresentanti tuttora viventi di questo phylum.

ticolari, tanto che i muscoli sono chiaramente visibili e il contenuto intestinale di alcuni esemplari può essere analizzato. *Ottoia* si cibava di due gruppi di animali forniti di conchiglia: i brachiopodi e gli iolitidi. Gli iolitidi, forse rappresentanti del phylum dei molluschi, avevano una conchiglia conica che poteva essere chiusa da un opercolo, quando l'animale si ritraeva completamente nel guscio. *Ottoia* non aveva denti abbastanza forti per spezzare e aprire la conchiglia, perciò inghiottiva gli iolitidi interi e digeriva le parti molli, mentre le conchiglie passavano intatte attraverso l'intestino del priapulide. Ma questi animali non erano il solo cibo di *Ottoia*; esiste un esemplare di priapulide che contiene nell'intestino i resti di un altro verme della medesima specie: questo sta a dimostrare che, come alcuni priapulidi attualmente viventi, *Ottoia* aveva forse abitudini cannibalistiche.

I parassitologi prendono in considerazione con estremo interesse un altro priapulide degli argilloscisti di Burgess, *Ancalagon*, forse antenato del gruppo attuale degli acantocefali, che sembrano essere stati parassiti per milioni d'anni. Questi vermi parassiti non hanno intestino e assorbono il nutrimento attraverso la parete corporea mentre sono alloggiati nell'intestino dell'ospite. Se immaginiamo di capovolgere per ipotesi la sequenza dell'evoluzione e di ridare a questi vermi gli organi necessari per un'esistenza libera, l'animale così ricostruito è notevolmente simile ad *Ancalagon*.

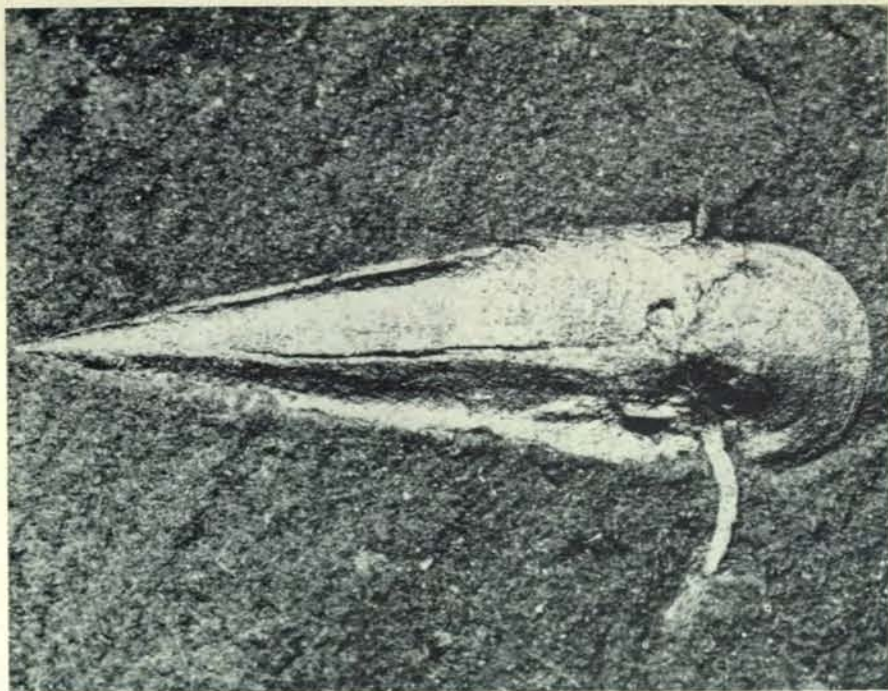
Due altri organismi, un tempo considerati anellidi policheti, sono *Wiwaxia* e *Pikaia*. Il corpo di *Wiwaxia* era ricoperto da grandi squame; le lunghe spine incurvate verso l'alto e verso l'esterno lungo il dorso evidentemente costituivano un'ottima protezione contro i predatori; che le spine fossero effettivamente usate come difesa è indicato dal fatto che in alcuni esemplari di *Wiwaxia* appaiono strappate. Abitatrice del fondale marino, *Wiwaxia* si nutriva «grattando» frammenti di cibo con un organo a lima; questa lima rassomigliava alla radula, la lingua dentellata e cornea di certi molluschi viventi. Si può considerare *Wiwaxia* un mollusco primitivo? Se è vera questa supposizione, i particolari dell'anatomia di questo animale, notevolmente conservati, potranno gettare una nuova luce sulle prime fasi dell'evoluzione di questo phylum di invertebrati che ha avuto tanto successo.

Che cosa si può dire invece di *Pikaia*, un tempo considerata un anellide policheте? Circa 30 esemplari ben conservati mostrano una struttura sporgente lungo il dorso dell'animale, che sembra essere una notocorda, il caratteristico organo scheletrico, fatto di tessuto simile a cartilagine, che dà il nome al phylum dei cordati. Oltre a questa importantissima caratteristica anatomica, si può notare che i blocchi muscolari in *Pikaia* formano un disegno a zig-zag, paragonabile a quello della muscolatura dell'anfiosso, cordato primitivo tuttora vivente, e dei pesci. Benché *Pikaia* differisca dall'anfiosso sotto parecchi aspetti importanti, in definitiva non si può assolutamente classificare come anellide, ma come cordato. Il meraviglioso stato di conservazione di questo organismo del Cambriano medio ne fa una pietra miliare nella storia del phylum cui appartengono tutti i vertebrati, compreso l'uomo. Esistono probabilmente esempi di cordati ancor più antichi, provenienti da formazioni del Cambriano inferiore della California e del Vermont, ma nessuno presenta una simile ricchezza di particolari.

Forse il problema più complicato presentato dalla fauna degli argilloscisti di Burgess è costituito dai 10 o più generi d'invertebrati che finora hanno resistito a tutti i tentativi di classificazione. Questi animali appaiono come gli unici rappresentanti noti di phyla la cui esistenza non era neppure stata sospettata precedentemente. La loro origine si perde nell'oscurità del Precambriano, quando iniziò la prima diversificazione dei metazoi. Un esempio delle insolite caratteristiche di questi organismi è dato da un genere avente un nome appropriato, *Hallucigenia*.

Questo animale si spingeva attraverso il fondale marino per mezzo di sette, appuntite paia di spine a forma di trampolo. Dalla superficie superiore dell'animale spuntavano sette tentacoli; all'estremità di ogni tentacolo c'era un paio di punte rinforzate. I tentacoli servivano forse a raccogliere cibo? Se sì, ogni tentacolo funzionava come una bocca individuale, con un collegamento diretto col canale alimentare dell'animale? Ci sono più domande che risposte, ma un indizio prezioso sul comportamento dell'animale è conservato in un reperto della raccolta della Harvard University, dove si possono vedere più di 15 esemplari di *Hallucigenia* associati a un grosso verme. Sembra quasi sicuro che questi strani animali, avendo scoperto la carcassa del verme, si siano raggruppati sopra di essa per cibarsene.

Confrontato con *Hallucigenia*, un altro singolare animale, *Opabinia*, sembra essere quasi normale. Aveva cinque occhi sistemati in vari punti della testa, cosicché probabilmente poteva evitare con facilità i predatori, quando nuotava vicino al fondale marino, regolando la propria direzione con una pinna caudale verticale. *Opabinia* si cibava di prede vive catturandole con un organo di presa che proiettava in avanti.

Fra questi strani invertebrati sono evidenti alcune soluzioni alternative ai problemi del disegno funzionale. Per esempio, per un animale vermiforme la cui cavità corporea sia riempita d'un fluido, la contrazione muscolare in una parte del corpo produce una distorsione della forma del resto del corpo. Negli anellidi il problema è stato risolto dividendo la cavità corporea in una serie di compartimenti a tenuta stagna. *Banffia*, un singolare organismo degli argilloscisti di Burgess, aveva sviluppato una soluzione alternativa. La metà anteriore del corpo, rinforzata da strutture rigide, era separata dalla metà posteriore, piuttosto simile a un sac-

Questo rappresentante del phylum dei molluschi, *Hyolithes*, aveva una conchiglia conica chiusa da un opercolo protettivo. Un priapulide scavatore, *Ottoia*, si cibava di questi molluschi, ma non era capace di spezzarne la conchiglia; perciò digeriva le parti molli e ne eliminava la conchiglia.

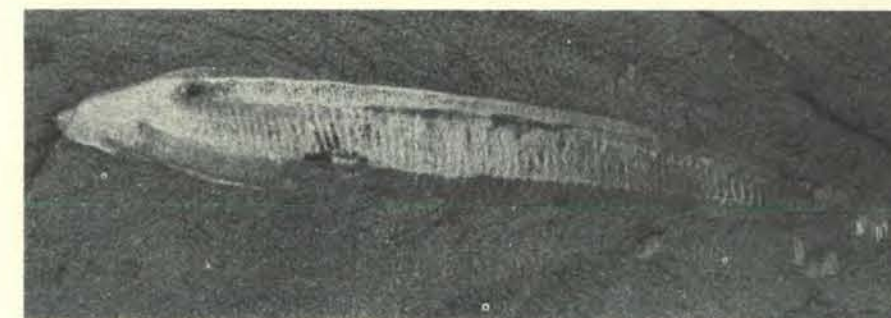
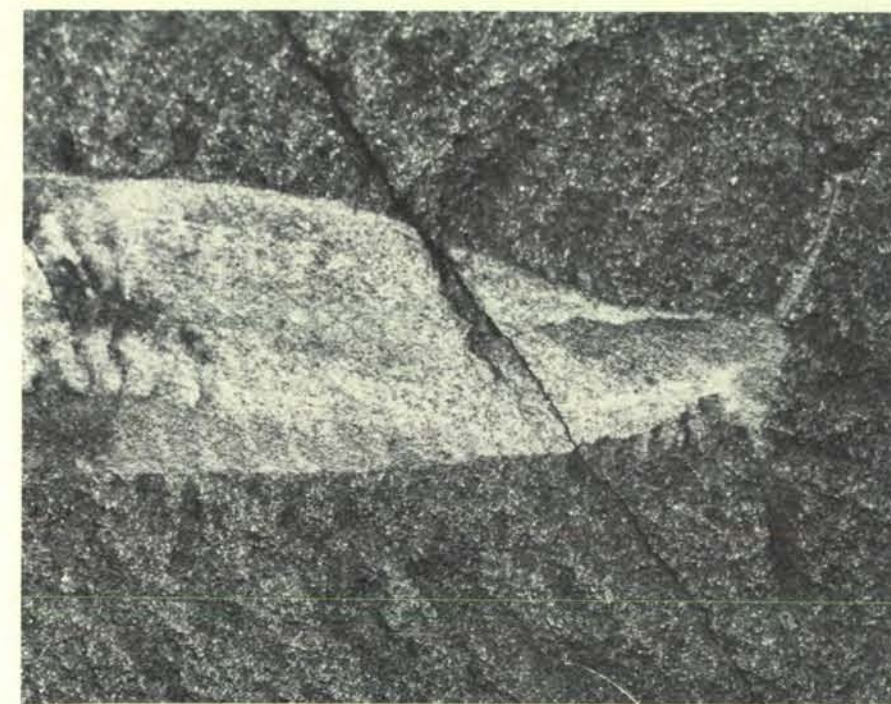
***Wiwaxia*, probabilmente un mollusco, con la sua corazza di grandi squame e la serie di lunghe spine protettive, fu classificata tra i policheti degli argilloscisti di Burgess. Ma il suo organo alimentare, a lima e simile a una radula di mollusco, indica l'appartenenza al phylum dei molluschi.**

co, da una costrizione. Questa costrizione sembra smorzasse le fluttuazioni idrostatiche prodotte dai muscoli locomotori della metà anteriore dell'animale, riducendo così al minimo le deformazioni della metà posteriore molle.

Alcuni rappresentanti di nuovi phyla sono stati conservati a dozzine. Altri, particolarmente gli abitatori liberamente natanti dei livelli acquatici superiori, sono rarissimi, poiché difficilmente potevano essere travolti dal limo che franava. Un animale di questo tipo è *Amiskwia*: a giudicare dalle sue pinne sporgenti, era probabilmente un nuotatore attivissimo. Un altro animale, *Nectocaris*, un predatore dal nuoto veloce, aveva occhi enormi ed evidentemente si spostava compiendo rapidi scarti laterali col corpo affusolato. Le pinne dorsali e ventrali, sporgenti e irrobustite da numerosi raggi, servivano a mantenere in equilibrio l'animale durante il nuoto.

I conodonti sono fossili enigmatici che somigliano a minuscoli denti; si trovano in formazioni che vanno dal tardo Precambriano al Triassico, un periodo lungo quasi 400 milioni d'anni. Benché sembrino denti, non possono aver avuto tale funzione, perché non mostrano segni di consunzione. Per lungo tempo l'appartenenza di questi conodonti a un qualche animale sconosciuto dal corpo molle e la funzione di tali organi sono state questioni irrisolte (si veda l'articolo *I conodonti* di Enrico Serpagli in «Le Scienze» n. 106, giugno 1977). Forse l'animale che conteneva i conodonti è *Odontogriphus*, un raro invertebrato pelagico conservato negli argilloscisti di Burgess, rappresentante di un phylum tuttora ignoto. L'apparato tentacolare dell'animale, usato per l'alimentazione, incorpora una serie di minuscoli oggetti conici che sembrano essere conodonti. Poiché i conodonti non possono aver avuto la funzione di denti, è stata avanzata l'ipotesi che fossero una specie di sostegno per i tentacoli alimentari. L'apparato alimentare di *Odontogriphus* e di animali simili a esso ha veramente prodotto tutti i conodonti così copiosamente diffusi durante tutto il Paleozoico e l'inizio del Mesozoico? Forse la risposta è positiva.

A mano a mano che si acquisiscono nuove conoscenze sulla fauna degli argilloscisti di Burgess, il quadro della vita nel Cambriano assume una nuova prospettiva, particolarmente nei riguardi della straordinaria evoluzione dei metazoi. Per esempio, la gamma di artropodi, con particolarissimi e differenti raggruppamenti di caratteristiche anatomiche, è talmente vasta che un solo phylum sembra troppo ristretto per poterla contenere interamente. La radiazione adattativa degli invertebrati del Cambriano può essere considerata come la risposta iniziale nei riguardi di una varietà assai ampia di nicchie ecologiche marine. Quindi molti animali del Cambriano appaiono come esperimenti pionieristici da parte di vari gruppi di metazoi, destinati a essere soppiantati a tempo opportuno da altri organismi meglio adattati. Dopo questa radiazione evolutiva del Cambriano, la tendenza sembra essere il successo e l'arricchimento nel numero di specie di relativamente pochi gruppi, a spese dell'estinzione di molti altri gruppi.

Una possibilità ulteriore viene suggerita dalla stessa fauna degli argilloscisti di Burgess. Alcuni gruppi predominanti durante il periodo cambriano, come i priapulidi, forse si sono trovati a mal partito di fronte ai competitori apparsi in seguito e sono riusciti a evitare l'estinzione migrando in nicchie marginali poco ambite o addirittura inutilizzabili da parte di altri metazoi. Una simile manifestazione di spostamento verso una nicchia marginale è la riduzione delle dimensioni corporee. Questa miniaturizzazione forse è proprio il sistema grazie al quale alcuni priapulidi sono riusciti a sopravvivere fino a oggi. Una via di scampo alternativa è la trasformazione in specie parassite; i priapulidi che appaiono aver dato origine agli acantocefali potrebbero essere un esempio di questa diversa via evolutiva. In ogni modo la fauna degli argilloscisti di Burgess fornisce uno splendido colpo d'occhio sulla dinamica dell'evoluzione durante questo breve intervallo del Cambriano medio e nel medesimo tempo costituisce un chiaro esempio di quanto siano impoveriti e distorti in genere i documenti fossili in nostro possesso. Lo studio di questi animali dal corpo molle getta una nuova luce su molti aspetti finora inimmaginabili della storia della vita.

Lungo una parte del dorso del cordato primitivo *Pikaia* appare una struttura rigida, la notocorda. La testa dell'animale, che si osserva nei particolari dell'illustrazione in basso, è a destra. La muscolatura ricorda nella sua forma e struttura quella dei pesci e dell'anfiosso. Nell'illustrazione delle pagine 76-77 è raffigurata una ricostruzione di questo animale, liberamente natante.

In questa fotografia si vede, ingrandita, l'estremità anteriore di *Pikaia*, con il paio di tentacoli sensoriali e, dietro, una fila di piccole appendici. Anche in altre formazioni più antiche del Cambriano esistono probabili resti di cordati, ma nessuno è paragonabile per ricchezza di particolari.

Il modello a sacche del confinamento dei quark

Benché si manifestino come entità reali, i quark non sono stati ancora osservati isolatamente, il che potrebbe essere attribuito al loro confinamento entro sacche analoghe alle bolle in un liquido

di Kenneth A. Johnson

Le particelle subatomiche vengono classificate in due ampie categorie: i leptoni (elettrone, muone, particella tau e rispettivi neutrini), insensibili all'interazione forte e verosimilmente privi di struttura interna, e gli adroni (tra cui il protone, il neutrone, il pione), che invece sono sensibili all'interazione forte e mostrano di possedere una struttura interna. La forza che unisce i neutroni e i protoni nel nucleo atomico è una manifestazione dell'interazione forte tra gli adroni. Tale interazione è una delle quattro forze della fisica; le altre sono la forza elettromagnetica, la forza debole e la forza gravitazionale.

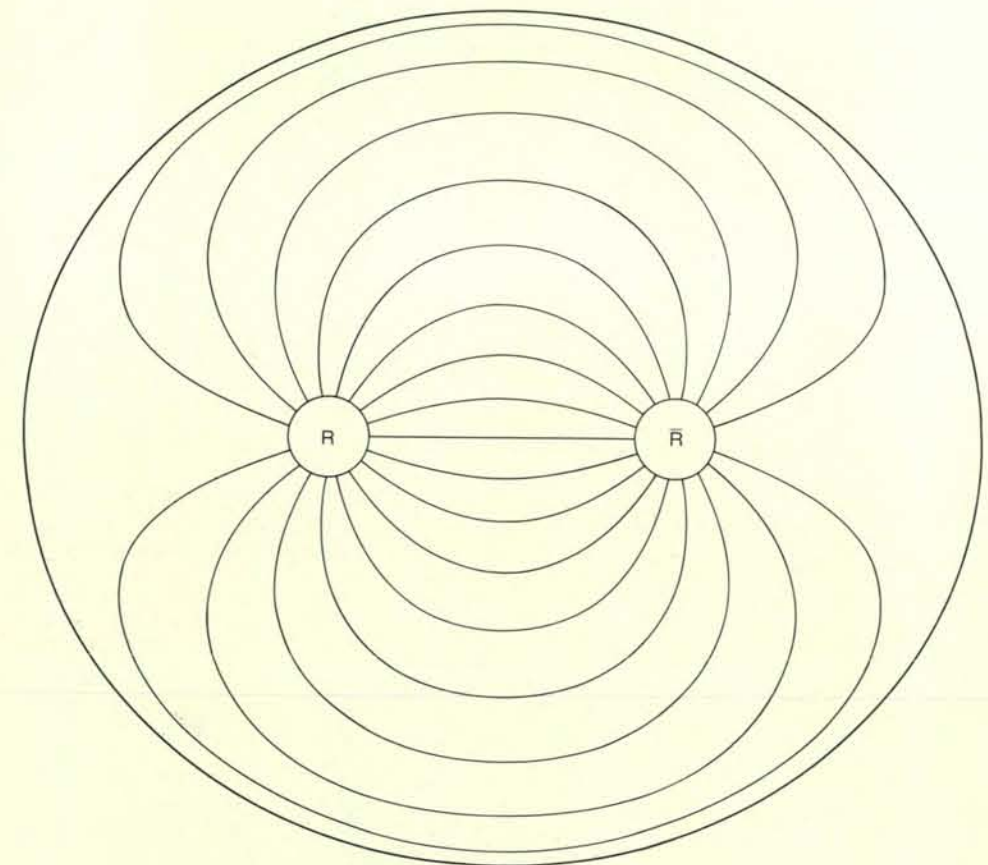
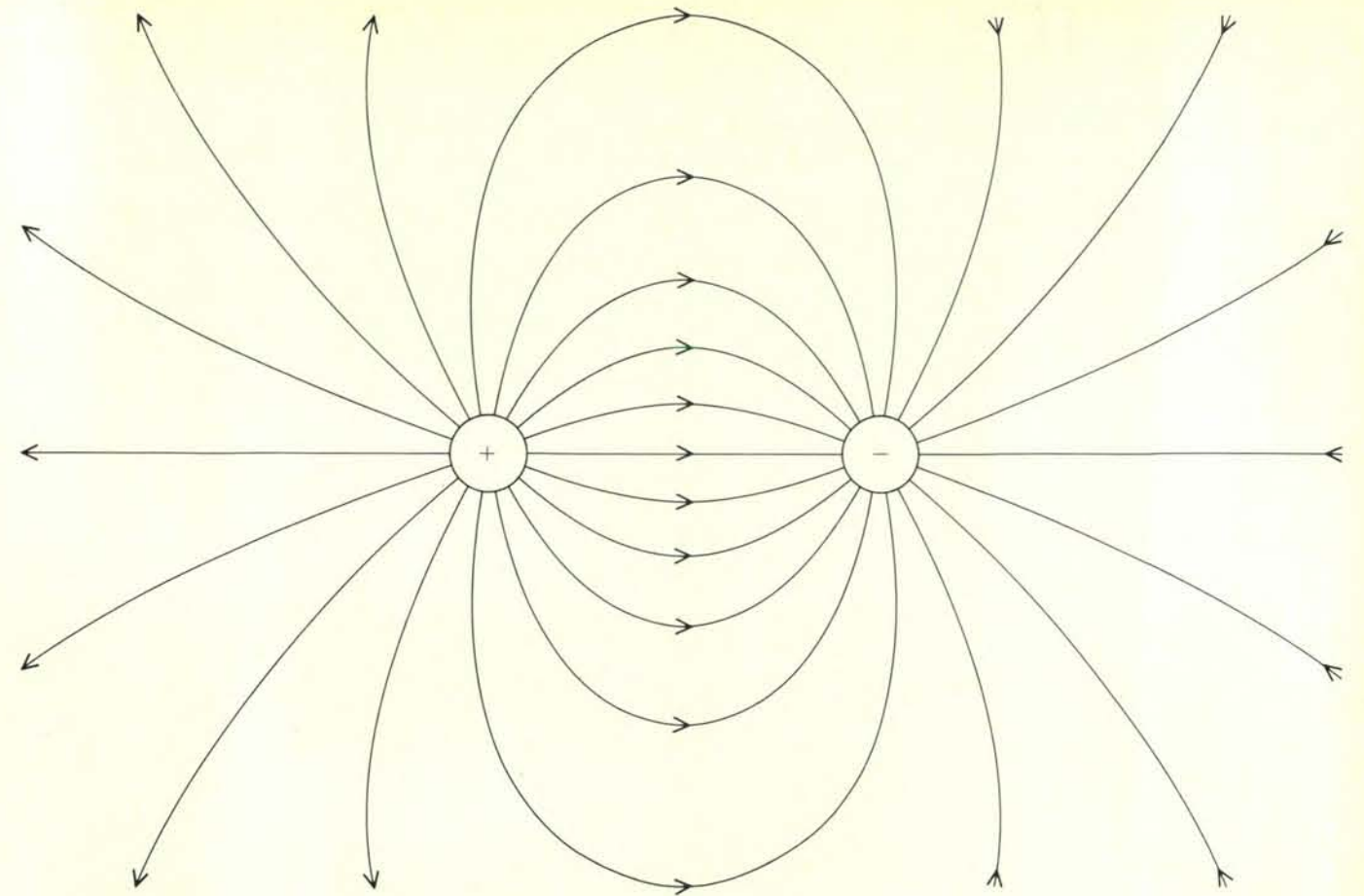
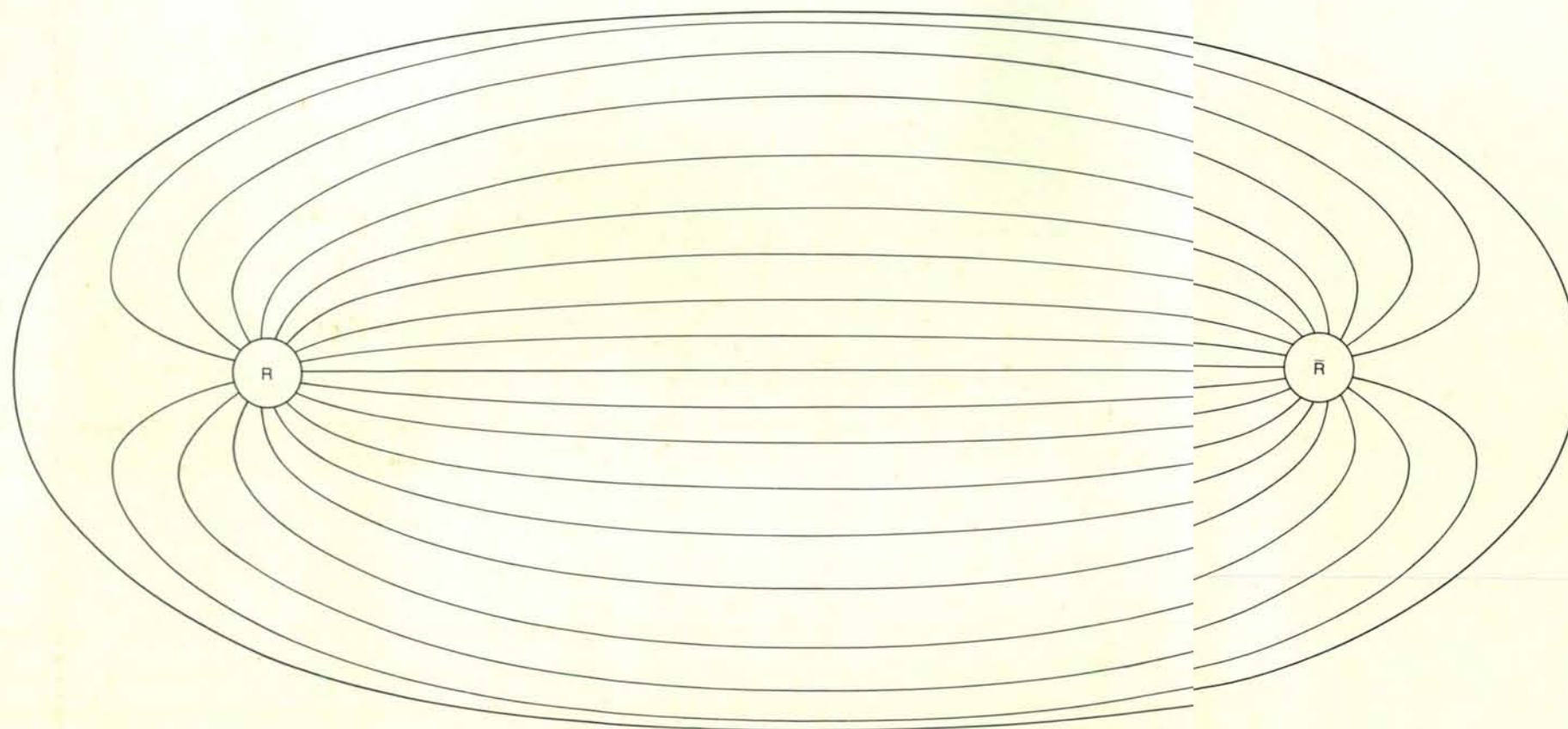
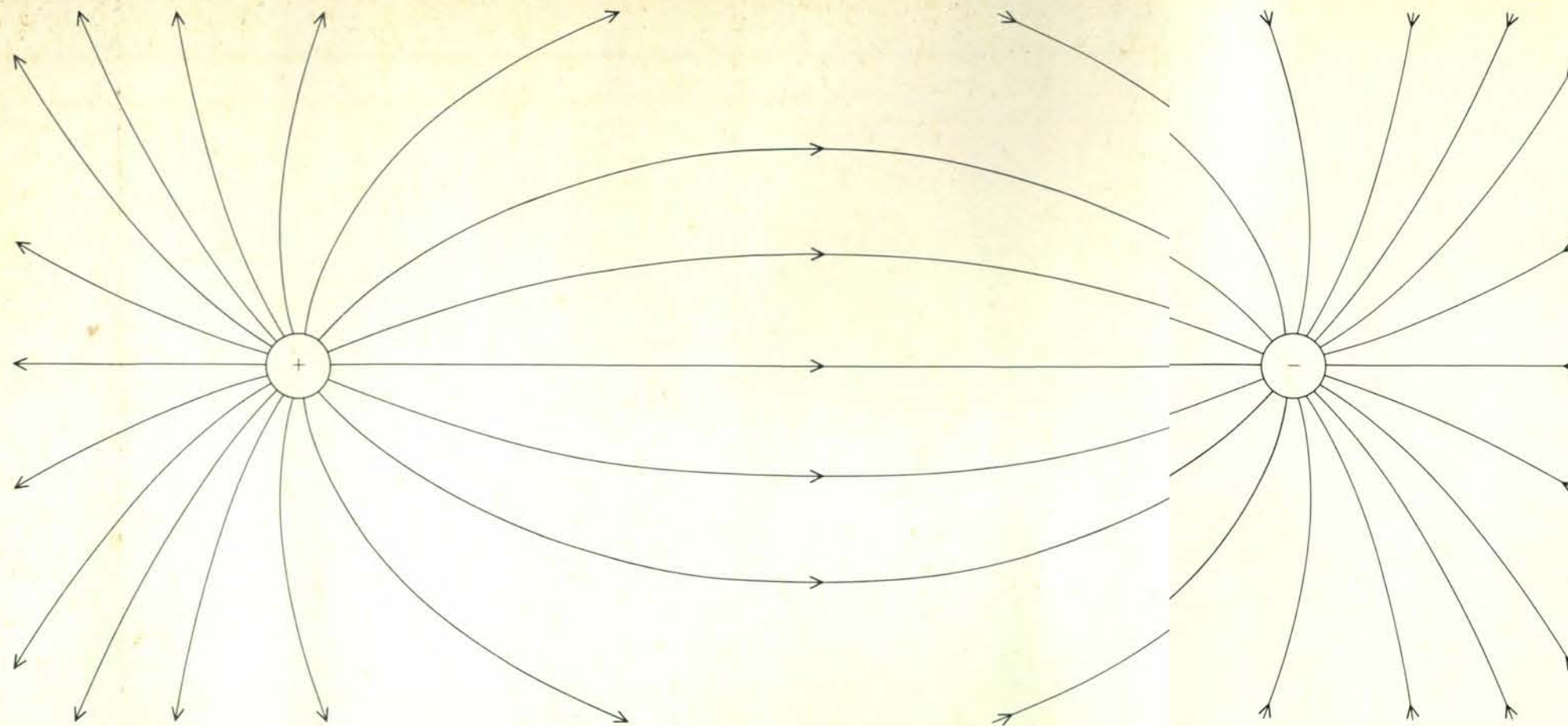
Nell'ultimo ventennio sono stati identificati oltre 200 tipi di adroni, che a loro volta sono stati suddivisi in due categorie, in relazione alla loro modalità di decadimento: i barioni, particelle che decadono essenzialmente in un protone, e i mesoni, che decadono sempre in leptoni e fotoni (quanti di energia elettromagnetica) o in coppie protone-antiprotone.

Tutti gli adroni sembrano costituiti dalla combinazione di entità denominate quark. Il modello a quark corrisponde a una straordinaria semplificazione della natura in quanto mediante cinque soli «sapori», o tipi, di quark (ma per lo più con tre soli), si riescono ora a interpretare le caratteristiche degli adroni. L'interazione forte tiene insieme i quark che costituiscono i barioni e i mesoni. Secondo il modello a quark, i barioni sono costituiti da tre quark e i mesoni da un'unica coppia quark-antiquark. L'enigma dei quark, ai quali sono state attribuite le etichette arbitrarie di «su», «giù», «strano», «incantato», «basso», risiede nel fatto che nonostante i molteplici tentativi non è stato ancora possibile osservarne alcuno isolatamente. Per spiegare questo fenomeno, io e il mio gruppo (Allen Chodos, Robert L. Jaffe, Charles B. Thorn e Victor F. Weisskopf) del Massachusetts Institute of Tecnology, abbiamo sviluppato un modello teorico secondo il quale i quark costitutivi di un adrone sono imprigionati in qualcosa di analogo a una sacca o a una bolla. Tale modello permette di chiarire molti aspetti della interazione forte che agisce tra i quark.

Quando nel 1963 Murray Gell-Mann e George Zweig, del California Institute of Technology, presentarono il modello a quark, questi non vennero accettati universalmente quali entità reali che sarebbe stato possibile osservare in laboratorio, ma vennero piuttosto considerati quali strumenti teorici che avrebbero consentito di mettere ordine nella complicata moltitudine degli adroni. I quark apparivano entità peculiari in quanto dotati di una carica elettrica pari a soltanto una frazione (+ 2/3 o – 1/3) della carica del protone, mentre la carica di ogni altra entità fisica nota è un multiplo intero di quella. Inoltre, sembrava che i quark violassero il principio di esclusione, che in meccanica quantistica rappresenta l'equivalente della nozione intuitiva secondo cui due oggetti non possono trovarsi nell'identica posizione allo stesso istante.

Alla fine degli anni sessanta, una serie di esperienze condotte dai gruppi di Jerome I. Friedman e Henry W. Kendall del MIT e di Richard E. Taylor dello Stanford Linear Accelerator Center (SLAC) portò la precisa conferma dell'importanza dei quark quali costituenti fondamentali di tutti gli adroni. In tali esperienze, consistenti nella diffusione anelastica di elettroni di alta energia da parte di protoni, la modalità della distribuzione degli elettroni diffusi indicava l'esistenza di particelle all'interno dei protoni. Tali particelle avevano un comportamento coincidente con quello postulato per i quark e quindi esse sono state identificate con i quark stessi. Le esperienze citate non hanno solo dimostrato la realtà fisica dei quark, ma anche la loro natura di entità puntiformi. Sembra quindi che tutti i componenti fondamentali della materia, sia quark che leptoni, si comportino come punti adimensionali.

Il fatto che un protone sia composto da tre quark spiega il perché della carica frazionaria di questi ultimi (il protone è infatti formato da due quark «su», ciascuno con carica + 2/3, e da un quark «giù», con carica – 1/3. Inoltre, l'introduzione di un nuovo numero quantico, cioè di una nuova proprietà fisica, denominato «colore», consente di dimostrare la compatibilità tra il comportamento dei quark e il principio di esclusione. Non essendo

Le linee di forza elettrica *(in alto)* generate da cariche elettriche di segno opposto sono curve nello spazio indicanti la direzione della forza elettrica cui sarebbe sottoposta una carica positiva posta sulle curve stesse. Il numero di linee in una data zona dello spazio indica l'intensità della forza: quanto maggiore il numero di linee, tanto più intensa la forza. Le linee del campo elettrico riempiono tutto lo spazio, ma lontano dalle cariche la loro densità si riduce e così pure la forza. Le linee di forza di «colore» *(in basso)* legano la coppia quark-antiquark di un tipico adrone, ossia di quelle particelle subnucleari soggette all'interazione denominata forte e che mostrano di possedere una struttura interna. Col termine colore si designa una proprietà del quark che non ha nulla a che fare con il colore visibile. Ciascun tipo di quark dispone di tre colori, «rosso», «blu», «giallo». Gli antiquark hanno gli anticolori «antirosso», «antiblu», «antigiallo». Tutti gli adroni sono «bianchi», ossia medie incolori dei tre colori e dei tre anticolori. Gli adroni a due quark sono costituiti da un quark di un colore, per esempio rosso *(R)*, e dall'antiquark del corrispondente anticolore, cioè antirosso *(R̄)*. Nel modello a sacche degli adroni, i quark sono confinati entro una sacca, o bolla, all'interno degli adroni, e quindi pure le linee del campo di colore sono confinate. Il fatto che le linee del campo di colore siano tra loro parallele, contribuisce a intensificare la forza agente tra le cariche, dal momento che tutte le linee esercitano la loro azione nella medesima direzione.

mai stati osservati allo stato libero, però, i quark restano tuttora oggetti enigmatici. È abbastanza agevole strappare un atomo da una molecola, un elettrone da un atomo o un protone da un nucleo atomico e i processi necessari allo scopo richiedono soltanto quantità di energia relativamente ridotte. Per contro, neppure con l'impiego di quantità di energia enormi è stato ancora possibile rimuovere un solo quark da un adrone.

Nei primi tempi alcuni fisici ritenevano che sarebbe stato possibile osservare quark liberi con l'ausilio delle energie superiori rese disponibili dal potenziamento degli acceleratori di particelle. Ciò non si è realizzato e attualmente si pensa che i quark siano definitivamente confinati all'interno degli adroni. Nel decennio scorso i fisici teorici hanno proposto diversi modelli del confinamento dei quark; il modello a sacche non è che uno di quelli. Esso sta acquisendo crescente favore perché è strettamente connesso con quella che la maggior parte dei fisici teorici considera la più soddisfacente fra le teorie fin qui formulate per le interazioni forti, ossia la cromodinamica quantistica, o dinamica del colore.

Intendo presentare il modello a sacche ricorrendo a una analogia suggerita da uno dei miei laureati, David Shalloway. Immaginiamo un mondo la cui atmosfera sia costituita di acqua anziché di aria; portando l'acqua a ebollizione spendendo una certa quantità di energia, si verrebbero a creare bolle di vapore acqueo, all'interno delle quali le molecole di vapore si comporterebbero, in prima approssimazione, come le particelle a interazione debole di un gas ideale. In questo ipotetico mondo le molecole di vapore potrebbero esistere solo all'interno delle bolle. Ritornando dalla bolla al liquido tali molecole cesserebbero di essere di vapore e perderebbero le proprietà di particelle di gas ideale.

Nel modello a sacche, il concetto di materia adronica è analogo. Secondo questo modello gli adroni sono sacche spaziali contenenti quark. All'interno delle sacche i quark si muovono liberamente e in modo indipendente, proprio come le molecole di vapore acqueo all'interno delle rispettive bolle. Se si tenta di rimuovere un quark dalla sacca, l'interazione forte di confinamento che agisce sul quark ha il sopravvento. Quindi i quark possono esistere solo entro le sacche che li contengono, analogamente alle molecole di vapore acqueo. In altri termini, gli adroni si comportano come sacche, o bolle di un vapore di quark ideale immerso in una atmosfera dello spazio ordinario entro la quale, invece, i quark non possono esistere.

Per portare l'acqua a ebollizione occorre spendere una certa quantità di calore latente; analogamente, per portare a ebollizione il «vuoto», ossia per convertire energia in materia, occorre spendere energia. Gli acceleratori di particelle possono essere considerati come ebollitori di vuoto. In essi una gran parte dell'energia cinetica delle particelle accelerate viene depositata su un piccolo bersaglio, ove per lo più viene spesa per la formazione di quark e anti-quark, che si combinano formando gli adroni osservabili nei prodotti di reazione. Nel modello a sacche una parte dell'energia va anche nel vuoto, interagendo con lo spazio ordinario in modo da creare le sacche che rinchiudono i quark. La quantità di energia necessaria per la creazione di una sacca è detta costante di sacca (B, da *bag*, che è il termine inglese per sacca) e vale circa 55 milioni di elettronvolt per fermi cubo (un fermi è uguale a 10^{-15} metri).

Se una particella è confinata in una regione di spazio di dimensioni Δx in ogni direzione, secondo il principio di indeterminazione della fisica quantistica la sua quantità di moto vale circa $\hbar/\Delta x$, ove $\hbar$ è la costante di Planck divisa per 2π. Ciò significa che i quark confinati in sacche di dimensioni finite si muovono con una quantità di moto dell'ordine di $\hbar/\Delta x$ e quindi esercitano una pressione sulla superficie interna delle sacche. In corrispondenza di un certo valore di equilibrio, la pressione dei quark sulla superficie interna bilancia la pressione di confinamento B che tiene la materia entro la sacca; da questo valore di equilibrio dipendono le dimensioni degli adroni. La pressione dello spazio ordinario in equilibrio con la pressione del vapore di quark che discende dal principio d'indeterminazione è qualcosa di analogo alla pressione dell'acqua in equilibrio con la pressione del vapor d'acqua nelle bolle.

Si potrebbe avere l'impressione che il problema del confinamento dei quark sia stato così banalmente risolto, con la semplice ipotesi dell'imprigionamento dei quark entro sacche, o bolle. Inoltre, una sacca contenente un unico quark sarebbe essa stessa un quark, anche se di una specie «fuori misura». Tale particella possiederebbe la stessa carica e altre proprietà di un singolo quark, e quindi il modello a sacche, come è stato fin qui illustrato, non spiegherebbe perché nessun quark è stato osservato isolatamente. Per interpretare il confinamento dei quark si deve introdurre la proprietà dinamica del colore. Ciascun sapore di quark possiede tre colori «rosso», «blu», «giallo». Gli antiquark hanno gli anticolori «antirosso», «antiblu», e «antigiallo». I colori dei quark non hanno niente a che fare con i colori visibili e il termine colore viene utilizzato solo perché le dif-

La separazione delle cariche influenza la forza agente tra esse. Separando due cariche elettriche opposte *(in alto)*, la curvatura delle linee di campo cresce. Le forze elettriche si riducono perché ha effetto solo la componente delle linee di campo nella direzione delle due cariche. La forza decresce secondo il quadrato della distanza tra le cariche (una legge di proporzionalità nota con il nome di legge di Coulomb). Separando due quark *(in basso)*, le linee non si curvano ma continuano ad agire nella stessa direzione. Dal parallelismo delle linee del campo di colore discende la costanza della forza di colore agente tra un quark e un antiquark.

ferenti modalità di combinazione tra quark nella meccanica quantistica ricordano quelle tra i colori visibili. Il modo migliore per immaginare tali colori (e pure i sapori) è di considerarli semplicemente come variabili supplementari associate ai quark. I quark hanno sapori e colori che possono anche assumere valori discreti diversi, esattamente come avviene per il momento angolare di spin di un quark, che ha valore $1/2\hbar$ e può assumere due orientazioni opposte.

Tutti gli adroni sono «bianchi», cioè medie incolori dei tre diversi colori. A grandi linee, i normali barioni sono costituiti di un quark rosso, uno blu e uno giallo, essendo il bianco la media di tali colori, mentre i mesoni consistono di un quark di un colore e di un antiquark del corrispondente anticolore. Una conferma della realtà del confinamento dei quark può essere tratta dalla dimostrazione che la caratteristica del colore non è mai osservabile, in quanto l'osservazione del colore implicherebbe quella di un quark libero. Attualmente si ritiene che l'interazione forte si basi sul colore dei quark proprio come l'interazione elettromagnetica si basa sulla carica elettrica. Le altre interazioni (elettromagnetica, debole e gravitazionale) sono invece tutte insensibili al colore.

È possibile illustrare più chiaramente gli effetti dinamici dei colori dei quark ricorrendo all'analogia con un concetto assai più antico, quello di carica elettrica. Michel Faraday rappresentò la carica come una sorgente di linee di forza elettrica. Una linea di forza è una curva nello spazio indicante la direzione della forza elettrica cui una carica positiva posta sulla curva sarebbe soggetta. Il numero di linee in una data zona di spazio indica l'intensità della forza: quanto maggiore il numero di linee, tanto maggiore la forza. Nella rappresentazione di Faraday, le linee di forza iniziano sulle cariche positive e terminano sulle cariche negative perché le cariche positive sono respinte da quelle dello stesso segno e attratte da quelle di segno opposto. Il numero di linee che originano o terminano su una carica dà una misura dell'intensità della carica stessa. Quindi, cariche intense generano forze elettriche intense.

I colori dei quark possono essere immaginati come particolari tipi di carica. Nell'elettrodinamica esiste un unico tipo di carica (che può essere positiva o negativa), ma nella cromodinamica ve ne sono diversi, ciascuno corrispondente approssimativamente alle combinazioni di colori permesse. L'azione della forza di colore non solo muove i quark ma ne modifica anche il colore, e quindi il campo di colore stesso deve essere colorato. (Un campo elettrico ha il solo effetto di muovere le cariche elettriche e non quello di modificarne il segno o l'intensità; questa caratteristica ha ricevuto generalizzazione matematica in una teoria dell'elettrodinamica sviluppata da C. N. Yang della State University di New York a Stony Brook e da Robert L. Mills dell'Ohio State University). Questa modalità d'interazione dei quark è coerente con i risultati delle esperienze in cui elettroni di alta energia subiscono diffusione anelastica da parte di protoni, esperienze analoghe a quelle che hanno rivelato che i protoni sono formati da quark. (Tale coerenza è stata evidenziata da David J. Gross e Franck Wilczeck della Princeton University e da H. David Politzer del Caltech). Fortunatamente non è indispensabile comprendere il meccanismo della trasmutazione dei colori per afferrare la relazione esistente tra il colore dei quark e il loro confinamento.

Proseguendo nell'analogia tra carica elettrica e carica di colore *(si vedano le illustrazioni da pagina 89 a pagina 91* è utile immaginare una linea di forza che inizia su una carica e termina su un'altra come a una specie di elastico che colleghi le cariche, perché la tensione lungo la linea di forza è parzialmente responsabile della forza di attrazione tra le cariche. Linee adiacenti esercitano una pressione l'una contro l'altra e perciò non si combinano per formare un'unica linea retta congiungente le cariche. La forza netta risultante fra le cariche è la somma delle forze associate a ciascuna linea. La curvatura delle linee cresce al crescere della separazione delle cariche. La forza viene a indebolirsi perché soltanto le componenti delle linee di campo nella direzione delle due cariche hanno effetto, e si riduce secondo il quadrato della distanza tra le cariche. Questa relazione di proporzionalità quadratica inversa è nota come legge di Coulomb.

		LEPTONI		QUARK	
PRIMA FAMIGLIA	PARTICELLA	e (ELETTRONE)	ν_e (NEUTRINO ELETTRONICO)	u (SU)	d (GIÙ)
	CARICA	−1	0	2/3	−1/3
	MASSA	$0,5 \times 10^{-3}$	0? ($< 0,6 \times 10^{-7}$)	$\sim 0,5 \times 10^{-2}$	$\sim 1 \times 10^{-2}$
	IPOTIZZATA	——	1929	1963	1963
	SCOPERTA	1898	1954	——	——
SECONDA FAMIGLIA	PARTICELLA	μ (MU)	ν_μ (NEUTRINO MUONICO)	c (INCANTATO)	s (STRANO)
	CARICA	−1	0	2/3	−1/3
	MASSA	0,11	0? ($< 0,57 \times 10^{-3}$)	~ 1,7	~ 0,3
	IPOTIZZATA	——	1957	1970	1963
	SCOPERTA	1936	1962	1974	1974
TERZA FAMIGLIA	PARTICELLA	τ (TAU)	ν_τ (NEUTRINO DELLA TAU)	t (ALTO O VERITÀ)	b (BASSO O BELLEZZA)
	CARICA	−1	0	2/3	−1/3
	MASSA	1,78	0? (< 0,25)	?	~ 5
	IPOTIZZATA	——	1975	1977	——
	SCOPERTA	1975	?	19??	1977
QUARTA FAMIGLIA?		?	?	?	?
QUINTA FAMIGLIA?		?	?	?	?
. . . ?		?	?	?	?

Le famiglie di particelle elementari sono costituite da due tipi di leptoni (particelle insensibili all'interazione forte e verosimilmente prive di struttura interna) e da due «sapori», o tipi, di quark, ciascuno dotato di tre diversi colori. Come i leptoni, i quark si comportano a modo di punti adimensionali. La carica si misura in unità di carica del protone e la massa si esprime in termini della sua energia equivalente, in miliardi di elettronvolt (GeV). Nessun quark è stato osservato isolatamente, ma è stato possibile individuarne i diversi sapori osservando gli adroni costituiti da due quark di ciascun sapore. Lo schema di classificazione in famiglie si fonda sul fatto che le modalità secondo cui i quark di una famiglia interagiscono fra di loro sono identiche a quelle che regolano le interazioni possibili fra i quark appartenenti a qualunque altra famiglia. I leptoni sono distribuiti nelle famiglie in ordine di massa crescente. Sono note fin qui tre famiglie di particelle, ma non esistono ostacoli, sul piano teorico, che escludano la possibilità di una scoperta, nel futuro, di ulteriori famiglie di particelle.

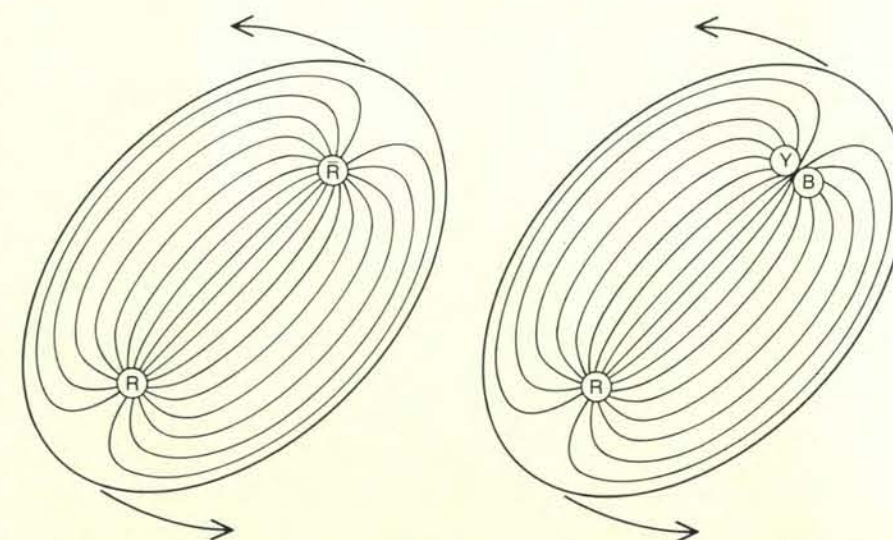
	NOME	SAPORI DI QUARK	SPIN	CARICA	MASSA (GeV)
MESONI	π^+	$u\bar{d}$	↑↓	+1	0,14
	K^+	$u\bar{s}$	↑↓	+1	0,50
	ρ^+	$u\bar{d}$	↑↑	+1	0,76
	ϕ	$s\bar{s}$	↑↑	0	1,04
	J/ψ	$c\bar{c}$	↑↑	0	3,10
	D^{0*}	$c\bar{u}$	↑↑	0	2,00
	S^*	$s\bar{s}$ O $(u\bar{u} + d\bar{d})$	↑↓↑↓	0	1.00
	Υ	$b\bar{b}$	↑↑	0	9,46
BARIONI	p^+	uud	↑↑↓	+1	0,938
	n	ddu	↑↑↓	0	0,940
	Δ^{++}	uuu	↑↑↑	+2	1,2
	Σ^+	uus	↑↑↓	+1	1,4
	Ω^-	sss	↑↑↑	−1	1,65

Nella tabella è riportato il contenuto in quark di alcuni adroni principali. I quark noti hanno cinque sapori: *u (up,* su), *d (down,* giù), *s (strange,* strano), *c (charmed,* incantato), *b (bottom,* basso). Gli antiquark hanno i simboli $\bar{u}, \bar{d}, \bar{s}, \bar{c}, \bar{b}$. Gli adroni sono stati divisi in due classi, in base alle loro modalità di decadimento. I barioni sono adroni che decadono essenzialmente in un protone; sono tutti costituiti da tre quark. I mesoni sono adroni che decadono interamente in fotoni e in leptoni o in coppie protone-antiprotone; per lo più i mesoni sono costituiti da un'unica coppia quark-antiquark. Le frecce nella colonna «spin» indicano la direzione di spin dei quark.

Una differenza sostanziale tra un normale campo elettrico e il campo di colore descritto dal modello a sacche consiste nel fatto che le linee di colore della forza che lega i quark non esistono ovunque ma sono confinate alle sacche. Alla superficie delle sacche, ove non esistono quark (e neppure, quindi, cariche di colore) le linee non possono avere né inizio né fine e, non potendo poi lasciare le sacche, esse devono essere tangenti alla superficie delle stesse. Sulla superficie la pressione verso l'esterno delle linee del campo di colore non è equilibrata da linee adiacenti, che sono assenti, ma dalla pressione B delle sacche di confinamento. Ne consegue che le linee di campo risultano fra loro parallele.

Il parallelismo tra le linee intensifica la forza agente tra le cariche perché tutte le linee hanno effetto nella stessa direzione. Poiché i quark (o cariche di colore) sono separati, le linee non curvano ma continuano ad avere effetto nella stessa direzione. Anche il numero delle linee resta invariato. È quindi una diretta conseguenza del parallelismo delle linee la costanza della forza tra un quark e un antiquark. L'intensità della forza non dipende dalla distanza tra i quark, fatta eccezione per il caso in cui tale distanza è estremamente ridotta.

La forza costante che si esercita tra quark di colore opposto è ciò che li tiene permanentemente insieme. L'intensità della forza, cioè la tensione netta delle linee del campo di colore che collegano le cariche è pari a circa 15 tonnellate. (Per contro, la forza elettrica che il protone esercita su un elettrone in un atomo di idrogeno è circa 10^{-11} tonnellate.) La forza di confinamento è dunque veramente forte. La costanza della forza che si esercita tra un quark e un antiquark indica che per separarli occorre ricorrere a una quantità di energia indefinitamente elevata, ciò che spiega il confinamento dei quark. Per esempio, per separare di un pollice (2,540 centimetri) il quark dall'antiquark in un mesone occorre una quantità di energia corrispondente a quella necessaria alla creazione di 10^{13} coppie protone-antiprotone.

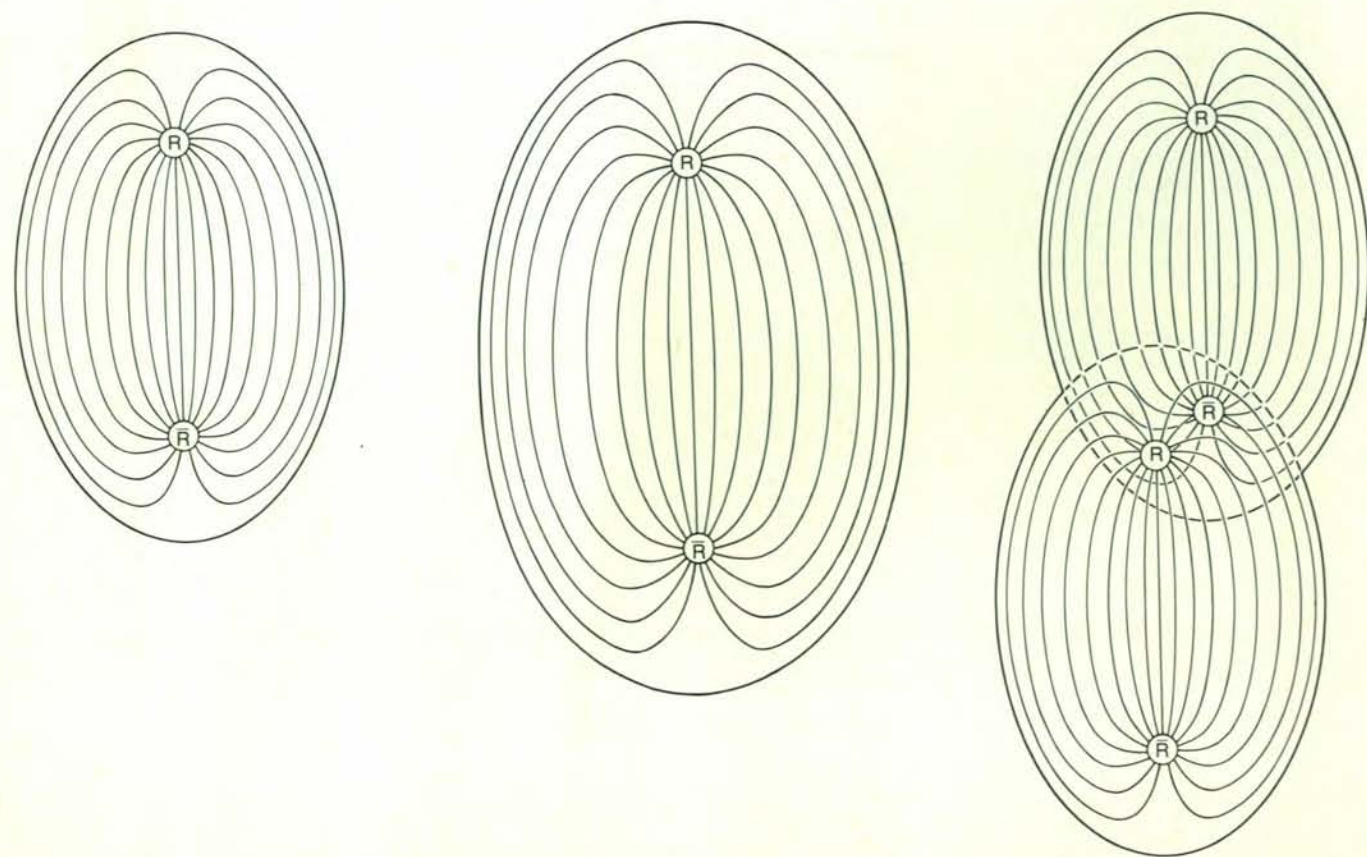
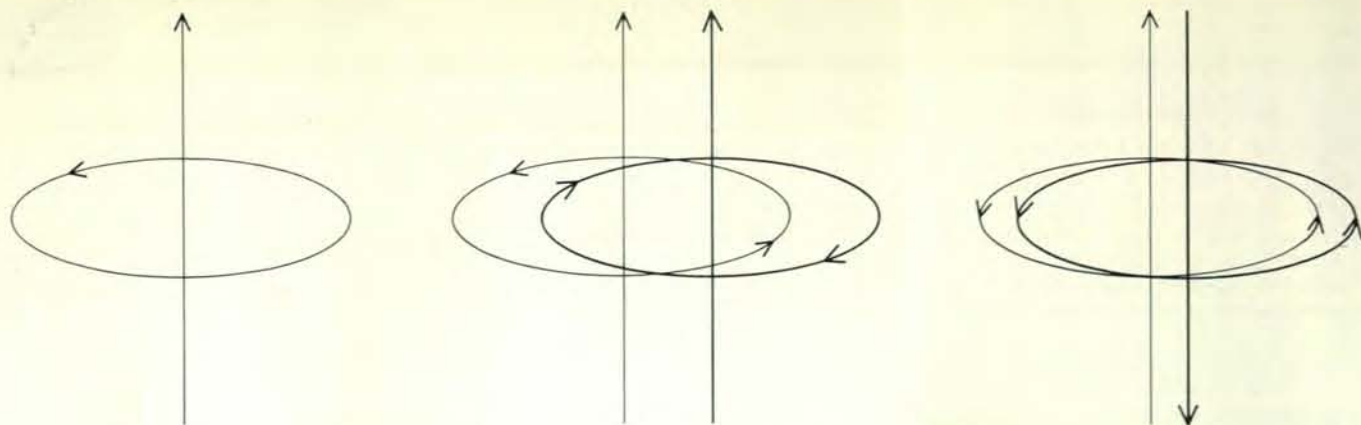
I soli aggregati di quark con massa finita sono quelli in cui le linee del campo di colore terminano tutte in un volume finito. Poiché le linee devono iniziare e terminare su cariche di colore, tali aggregati devono essere combinazioni incolori di quark. In altri termini, e in accordo con le osservazioni sperimentali, tutti gli adroni sono bianchi. Il problema del confinamento dei quark è stato risolto riconducendolo a quello del confinamento del colore. Il modello a sacche implica appunto il confinamento dei colori. Le modalità di combinazione dei colori spiegano perché i barioni consistono di soli tre quark e il mesone di una singola coppia quark-antiquark. L'aggregato bianco più semplice è costituito da un quark e da un antiquark di colori opposti. Quello immediatamen-

La relazione tra momento angolare e massa è identica per i mesoni e per i barioni: questo fatto viene spiegato dal modello a sacche. Se un quark e un antiquark in un mesone sono messi in rotazione *(a sinistra)*, essi si allontanano reciprocamente come due pattinatori che pattinano in cerchio tenendo gli estremi di una fune elastica, con la differenza che per i quark la fune è data dalle linee del campo di colore colleganti le cariche di colore opposte. Se il sistema ruota, il momento angolare cresce, le linee di colore si tendono e nel campo di colore rappresentato dalle linee si immette energia, con un incremento della massa del sistema. La relazione dipende solo dalla differenza di colore tra i quark agli estremi della particella. Poiché il barione è bianco *(a destra)*, i due quark (giallo e blu) a un estremo hanno un colore risultante (antirosso) opposto al colore (rosso) del quark all'altro estremo. Quindi i quark hanno lo stesso colore dell'antiquark del mesone.

te successivo in ordine di complessità crescente è costituito da un quark rosso, uno blu e uno giallo. Non è possibile avere una combinazione incolore di quark se la differenza tra il numero di quark e quello degli antiquark non è zero o un multiplo di tre.

La cromodinamica, insieme con il modello a sacche, può spiegare molti aspetti del comportamento degli adroni che prima apparivano paradossali. Per esempio, se un quark e un antiquark vengono posti in rotazione, essi si allontanano reciprocamente come due pattinatori che pattinano in cerchio tenendo gli estremi di una fune elastica, con la differenza che per i quark la fune corrisponde alle linee di colore colleganti le cariche di colore opposte. Quando il sistema è in rotazione il suo momento angolare cresce, le linee di colore si tendono ed entro il campo di colore che le linee rappresentano viene a immettersi energia. Ne consegue un incremento della massa del sistema. La relazione prevista dal modello tra momento angolare di spin e massa ha avuto verifica sperimentale per tutti gli adroni *(si veda l'illustrazione in basso nella pagina precedente)*. La scoperta sperimentale di tale relazione è stata uno dei punti di forza di un altro modello a corde, in cui due quark non possono essere separati perché sono legati da una «corda». Il modello a sacche fornisce una spiegazione fisica della natura di tale corda.

La formulazione della relazione tra massa e momento angolare derivata in base al modello a sacche indica perché per i mesoni e i barioni valga la stessa relazione. La relazione infatti dipende solo dalla carica di colore dei quark agli estremi della particella e poiché il barione è bianco, i quark a un suo estremo presentano una combinazione di colore opposta al colore del quark all'altro estremo. In altri termini, essi devono avere quale risultante lo stesso colore netto dell'antiquark nel mesone.

Anche la creazione di adroni si può spiegare mediante il concetto di colore. Allo stesso modo dell'elettrone e del positone (elettrone di carica positiva), anche un quark e un antiquark possono essere creati come coppia. Il relativo processo richiede la presenza di un «campo polarizzatore» che separi le cariche di colore di una coppia quark-antiquark in un mesone. Una nuova coppia appare spontaneamente nel campo di colore del quark e dell'antiquark del mesone. Le linee di colore legano il quark del mesone all'antiquark di nuova creazione e l'antiquark del mesone al nuovo quark. Entrambe le coppie quark-antiquark sono incolori, e quindi le linee di colore non intervengono. Le coppie si possono separare, in quanto tra loro non agisce una forza costante; in altri termini, l'interazione forte può trasformare un mesone in altri due, per esempio un mesone ϱ in due pioni (mesoni π).

Il concetto di linea di campo di colore è stato usato con successo anche per descrivere l'auto-annichilazione di un elettrone e di un positone a energie elevatissime. Se le particelle sono dotate di energia sufficientemente elevata, possono generare un campo elettromagnetico di intensità tale da polarizzare il vuoto e da creare una coppia quark-antiquark di cariche opposte. Se la coppia è costituita di quark con massa piccola, questi si allontanano con la stessa velocità con cui si incontrano l'elettrone e il positone ma non necessariamente lungo lo stesso asse. Ovviamente il quark e l'antiquark sono collegati da linee del campo di colore. Mentre il quark e l'antiquark si allontanano rapidamente, l'intenso campo di colore crea coppie di quark colorati costitutivi di mesoni. La pressione delle linee del campo di colore confina la coppia quark-antiquark divergente entro una zona tubolare di spazio dello spessore di circa 10^{-15} metri. Per il principio d'inde-

La trasformazione di un mesone *(a sinistra)* in due mesoni implica un «campo polarizzatore» che separi le cariche di colore della coppia di quark del mesone *(al centro)*. Una nuova coppia quark-antiquark appare spontaneamente nel campo polarizzatore tra il quark e l'antiquark del mesone *(a destra)*. Le linee di colore legano i quark del mesone all'antiquark di nuova creazione e l'antiquark del mesone al nuovo quark. Tali coppie quark-antiquark sono incolori, e quindi le linee di colore non le raggiungono; esse possono separarsi e formare due mesoni, poiché tra loro non agisce una forza costante, come invece accade per il quark e l'antiquark in ciascuna coppia.

Il campo magnetico di colore è generato da un quark in moto (carica di colore), esattamente come un campo magnetico ordinario è generato da cariche elettriche in moto. I quark hanno momento angolare di spin; sono quindi sempre in rotazione e possono essere trattati come un anello di corrente in cui una determinata carica di colore circoli (in senso antiorario per lo spin rivolto verso l'alto) su un piano ad angolo retto rispetto alla direzione dello spin del quark *(a sinistra)*. Se gli spin sono paralleli, i quark nel mesone sono di colore opposto; quindi le correnti di colore hanno direzioni opposte e conseguentemente i quark si respingono magneticamente l'un l'altro *(al centro)*. Se gli spin sono antiparalleli, le correnti di colore hanno la stessa direzione, per cui la forza magnetica di colore attira i quark l'uno verso l'altro *(a destra)*. Da questo effetto dipende il fatto che gli adroni con spin nella stessa direzione possiedano massa maggiore rispetto a quelli con spin opposti.

terminazione, la componente della quantità di moto della coppia in direzione normale al tubo risulta equivalente solo ad alcune centinaia di milioni di elettronvolt. Poiché il quark e l'antiquark divergono rapidamente, la componente della quantità di moto lungo l'asse determinato dalla loro direzione di movimento è molto più elevata. Ne consegue che gli adroni risultanti si muovono prevalentemente lungo tale direzione. Questi sciami di adroni sono stati osservati recentemente allo SLAC e in altri laboratori.

L'elettricità e il magnetismo sono fenomeni strettamente connessi: basti pensare che le cariche elettriche in moto generano campi magnetici. Poiché la carica di colore è stata introdotta per analogia con la carica elettrica, si può forse anche parlare di una specie di magnetismo di colore (cromomagnetismo). Tale tipo di magnetismo esiste davvero, e uno dei suoi effetti è corrisposto a uno dei primi successi della cromodinamica. Questo effetto è stato scoperto indipendentemente da Howard Georgi, Alvaro De Rújula, Sheldon L. Glashow della Harvard University e da un gruppo di noi al MIT. È ben noto sperimentalmente che le correnti che fluiscono nella stessa direzione lungo due fili paralleli generano una forza magnetica che induce attrazione tra i fili stessi; se invece le stesse correnti hanno tra loro direzioni opposte, la forza magnetica che viene generata induce repulsione tra i fili.

Essendo dotati di momento angolare di spin, i quark si comportano come cariche di colore in moto e quindi generano un campo magnetico di colore. Un quark in rotazione può essere immaginato come un anello di corrente di colore in cui una determinata carica di colore circola su un piano posto ad angolo retto rispetto alla direzione dello spin del quark. In un mesone i quark sono di colori opposti: se quindi i relativi spin sono paralleli, le correnti di colore hanno direzioni opposte e pertanto i quark si respingono magneticamente l'un l'altro. D'altra parte, se gli spin sono antiparalleli, le correnti di colore hanno la stessa direzione, con il risultato che la forza magnetica di colore spinge i quark l'uno verso l'altro. Da ciò dipende la maggiore massa delle particelle i cui quark hanno spin nella stessa direzione, rispetto a quelle con spin in direzioni opposte. Questa conseguenza della cromodinamica è stata osservata per l'intera famiglia degli adroni. Per esempio, il mesone ρ, che ha paralleli gli spin dei quark, è più pesante del mesone π, nel quale invece gli spin dei quark costitutivi sono antiparalleli.

Il fatto che il nucleo atomico sia costituito da un insieme di neutroni e di protoni, anziché direttamente di quark, può dipendere molto probabilmente dall'intensità delle forze magnetiche di colore. L'energia di un aggregato di quark di tipo «su» e «giù» è inferiore qundo tali quark dapprima si associano in masse incolori (neutroni e protoni), le quali poi si vincolano reciprocamente (anche se in modo relativamente debole) formando un nucleo atomico. Ciò significa che se la materia nucleare venisse in qualche modo compressa almeno fino a raddoppiare la propria densità, tutti i nucleoni tenderebbero a perdere la propria identità confondendosi in un fluido di quark. Il passaggio dalla materia nucleare alla materia di quark probabilmente non è una transizione di fase, del tipo che s'incontra nella trasformazione di gas in liquido al ridursi della temperatura; si tratterebbe invece di una transizione più graduale, simile alla trasformazione di un gas monoatomico in gas biatomico al diminuire della temperatura.

La compressione della materia nucleare in laboratorio presenterebbe enormi difficoltà; eppure sembra che la materia nucleare venga compressa da enormi forze gravitazionali nelle stelle di neutroni, ossia nei resti freddi dell'esplosione delle supernove. Quando la densità della materia neutronica compressa supera la densità dei quark nel neutrone, i neutroni stessi fondono in un fluido di quark. Pertanto, la zona centrale densa di una stella di neutroni può realmente consistere di quark.

Questa possibilità ha notevole effetto per la stabilità delle stelle di neutroni. La materia costitutiva della stella è compressa dalla massa gravitazionale, che per lo più è massa neutronica. La resistenza alla compressione deriva dall'interazione repulsiva dei neutroni e dall'energia cinetica dei neutroni freddi. Si ritiene che dall'energia cinetica dei quark vincolati dipenda la parte preponderante della massa neutronica, e che quindi la pressione crescerebbe enormemente se i neutroni si fondessero tra loro e liberassero i quark. Questo effetto produrrebbe un sensibile incremento della massa permessa di una stella di neutroni stabile. Sono in corso studi miranti a individuare le proprietà della materia di quark che consentirebbero di effettuare un calcolo della massa permessa.

Dal momento che con il modello del campo di colore si possono descrivere così tanti aspetti delle proprietà e delle interazioni degli adroni, i fisici nucleari sperano che il confinamento dei quark non sia altro che una necessaria conseguenza della cromodinamica. È possibile che il modello a sacche degli adroni discenda dalla cromodinamica? La risposta è positiva. Per comprendere come stiano le cose, si consideri una possibile analogia con la superconduttività (la proprietà consistente nell'assenza di resistenza elettrica nei metalli a temperature prossime allo zero assoluto). Una forza magnetica applicata non può penetrare entro un corpo superconduttore; infatti, il campo magnetico induce correnti elettroniche entro un sottile strato superficiale del corpo stesso; tali correnti a loro volta creano un campo magnetico che si oppone al campo applicato con un'intensità esattamente sufficiente a renderlo nullo all'interno del corpo. Tuttavia, in talune condizioni, il campo magnetico applicato può penetrare nel corpo creandovi una zona di conduttività normale, all'interno della quale le linee di flusso magnetico possono essere intrappolate. Questo fenomeno è denominato effetto Meissner.

Nella cromodinamica può esistere un corrispondente dell'effetto Meissner. Forse il vuoto nella cromodinamica quantistica respinge le linee del campo di colore esattamente come un superconduttore respinge le linee di campo magnetico. In tal caso i campi di colore potrebbero esistere solo laddove da essi fosse stata spesa energia per creare zone di «vuoto normale» (analoghe a quelle di conduttività normale dell'effetto Meissner) entro il vuoto circostante. La proprietà che il vuoto circostante avrebbe, cioè quella di espellere il calore assolverebbe la stessa funzione della pressione della sacca nel modello a sacche, mantenendo cioè confinate le linee del campo di colore all'interno degli adroni. A sostegno di questo modello di «vuoto duale», sono stati eseguiti approfonditi calcoli da parte di Charles G. Gross, Curtis G. Callan e R. Dashen della Princeton University. Un altro modello di questo genere è stato studiato da Holger B. Nielsen e M. Minomiya dell'Istituto Niels Bohr (Danimarca). Su tali modelli si fondano le ricerche in corso attualmente per chiarire il meccanismo del confinamento dei quark.

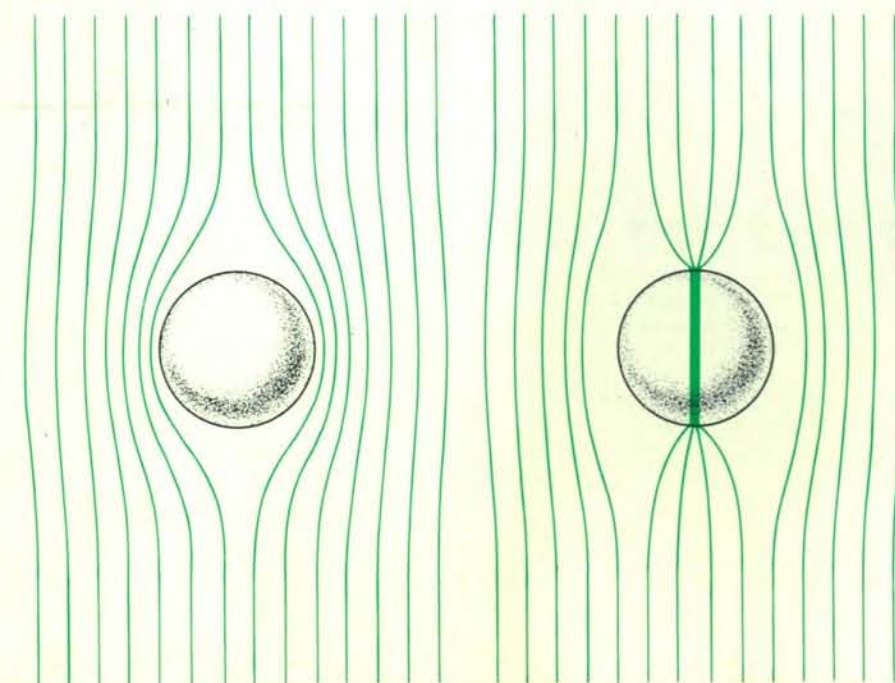
Mediante la cromodinamica quantistica e il modello a sacca si è in buona misura interpretato il meccanismo dell'isolamento permanente dei quark all'interno degli adroni, e quindi si è data una spiegazione dei vari aspetti dell'interazione forte tra i quark. La validità completa dell'interpretazione dipenderà dalla sua conferma da ottenersi mediante calcoli più raffinati e dettagliati delle proprietà degli adroni. Il fenomeno del confinamento dei quark è forse l'aspetto più paradossale del modello a quark, e una sua interpretazione definitiva varrà a eliminare i residui dubbi sull'ipotesi, peraltro fortunata, dei quark. È giunto il momento di riconsiderare la rappresentazione delle particelle subatomiche alla luce del modello a quark. Questi vengono considerati i costituenti elementari della materia poiché sembrano interagire come punti adimensionali, ma sono veramente elementari? Nella storia della scienza si riscontrano molti esempi di entità fisiche ritenute originariamente elementari ma in seguito rivelatesi non tali. Gli atomi, i nuclei atomici, i protoni e i neutroni sono stati tutti considerati indivisibili in vari momenti del nostro secolo, fino a che ulteriori esperienze ne hanno messo in evidenza i rispettivi costituenti. È possibile che si verifichi lo stesso per i quark e per i leptoni, che pure oggi sono ritenuti elementari?

La meccanica quantistica e la teoria della relatività non escludono affatto la possibilità che i quark o i leptoni abbiano una struttura interna. I quark e i leptoni sembrano avere dimensioni non superiori a 10^{-17} metri. Per l'elettrone, tale limite dimensionale superiore è stato definito sperimentalmente. Sulla base del principio d'indeterminazione, è possibile calcolare la quantità di moto di un'ipotetica entità costitutiva confinata in una zona di spazio di diametro 10^{-17} metri. Secondo la teoria della relatività, l'energia cinetica di tale entità sarebbe maggiore del prodotto della sua quantità di moto per la velocità della luce. L'energia cinetica dell'entità confinata in una zona di 10^{-17} metri di diametro farebbe crescere la massa del quark o del leptone di cui facesse parte; l'entità di energia minima darebbe un contributo di massa pari ad almeno 20 volte la massa del protone. È pertanto estremamente improbabile, per esempio, che l'elettrone abbia dei costituenti elementari. La presenza di tali costituenti implicherebbe infatti il verificarsi di un annullamento casuale e praticamente perfetto delle rispettive energie di legame da parte delle corrispondenti energie cinetiche. Tali eventualità si verificano raramente in fisica ed è quindi estremamente improbabile che gli elettroni (come pure gli altri leptoni e i quark) abbiano una struttura composta.

Se veramente i quark e i leptoni sono elementari, la natura non dovrebbe continuare a complicare l'immagine della materia creando nuovi tipi di particelle sempre più fondamentali. Sembra invece che l'immagine si stia complicando per la presenza di nuovi sapori di quark e di nuove specie di leptoni. I quark e i leptoni noti si possono classificare raggruppandoli in famiglie. Ciascuna famiglia è costituita da quattro particelle con carica elettrica globale nulla, ossia due leptoni (una particella e il suo corrispondente neutrino) e due sapori di quark (nei tre diversi colori, rosso, blu e giallo). Finora sono state individuate tre di tali famiglie. Alla prima appartengono l'elettrone, il neutrino elettronico, il quark «su» e il quark «giù». Alla seconda il muone, il neutrino muonico, il quark «incantato» e il quark «strano». Alla terza la particella tau, il neutrino della tau, il quark «basso» e un ipotetico quark «alto», non ancora rivelato ma previsto teoricamente.

I costituenti della prima famiglia svolgono il ruolo principale nei fenomeni naturali. Le altre particelle elementari, create essenzialmente negli acceleratori, hanno un'esistenza effimera e sulla materia ordinaria influiscono solo mediante effetti estremamente tenui. Esse decadono per interazione debole nei costituenti della prima famiglia. Questo schema di classificazione si fonda sul fatto che le quattro particelle in una data famiglia interagiscono reciprocamente allo stesso modo delle quattro di qualsiasi altra famiglia. Per esempio, in una famiglia un

	SPIN PARALLELI			SPIN ANTIPARALLELI		
MOMENTO ANGOLARE ORBITALE	PARTICELLA	CONTENUTO IN QUARK	MASSA (GeV)	PARTICELLA	CONTENUTO IN QUARK	MASSA (GeV)
0	Δ^+	uud	1,23	p	uud	0,94
0	Σ^{+*}	uus	1,38	Σ^+	uus	1,19
0	Ξ^{-*}	dss	1,54	Ξ^-	dss	1,32
0	ρ^+	$u\bar{d}$	0,77	π^+	$u\bar{d}$	0,14
1	A_2^+	$u\bar{d}$	1,31	B^+	$u\bar{d}$	1,23
2	g^+	$u\bar{d}$	1,69	A_3	$u\bar{d}$	1,64
0	K^{-*}	$u\bar{s}$	0,89	K^-	$u\bar{s}$	0,49
1	K^{-**}	$u\bar{s}$	1,42	"Q"	$u\bar{s}$	1,3
2	K^{-***}	$u\bar{s}$	1,78	L^-	$u\bar{s}$	1,77

Sono qui raffigurati per vari adroni gli effetti di massa del magnetismo di colore. La tabella indica chiaramente che, a parità di ogni altra condizione, una particella con spin dei relativi quark paralleli ha massa maggiore di una con spin antiparalleli. Il momento angolare orbitale è espresso in unità della costante di Planck divisa per 2π. Ruotando, i quark vengono allontanati reciprocamente dalla forza centrifuga cui sono soggetti nel sistema di riferimento in rotazione; la loro distanza viene a crescere con il crescere del ritmo della rotazione. Al crescere della distanza si affievolisce l'interazione magnetica di colore, il che significa che aumentando il momento angolare (e quindi il ritmo della rotazione) le masse negli stati paralleli e antiparalleli convergono.

Ricorrendo all'analogia con le proprietà di un superconduttore (un metallo privo di resistenza elettrica alle temperature prossime allo zero assoluto), si può mostrare come il confinamento dei quark sia una conseguenza della cromodinamica, o dinamica del colore. Un campo magnetico applicato non può penetrare entro un corpo superconduttore *(a sinistra)*. Tale campo induce correnti elettroniche in un sottile strato superficiale del corpo, le quali a loro volta generano un campo magnetico opposto e di intensità esattamente sufficiente ad annullare entro il corpo il campo applicato. Il campo magnetico applicato può penetrare nel corpo in un unico modo, e cioè in seguito alla creazione di una zona di conduttività normale, entro la quale le linee di flusso magnetico possono essere intrappolate *(a destra)*. Nella cromodinamica quantistica il vuoto può respingere le linee del campo di colore, proprio come un superconduttore respinge le linee del campo magnetico. I campi di colore possono esistere solo in regioni ove in precedenza abbiano speso energia per creare, entro il vuoto circostante, zone di «vuoto normale» analoghe alle zone di conduttività normale del superconduttore. La caratteristica del vuoto circostante rappresentata dalla repulsione del colore svolgerebbe la stessa funzione che svolge la sacca nel modello a sacche.

sapore di quark può essere trasformato nell'altro mediante l'interazione debole. (È quanto si verifica nella normale radioattività, ove un quark «su» si trasforma in un quark «giù» e viceversa. Allo stesso modo, un leptone si può trasformare in un altro leptone. I quark possono anche interagire incrociando linee di famiglie, benché tale interazione sia assai più debole. Non sono mai stati individuati casi di leptoni che incrocino linee di famiglie, né leptoni che si siano trasformati in quark o quark che si siano trasformati in leptoni.

Tutti i costituenti della prima famiglia sono stati individuati alla fine degli anni sessanta, con la rivelazione dei quark «su» e «giù» all'interno del protone e del neutrone. Le particelle della seconda famiglia sono state scoperte tutte tra il 1936 e il 1974. La rivelazione nel 1974 di un quark dotato di incanto, ultimo componente della seconda famiglia, ha rappresentato un sostegno decisivo per il modello a quark. La necessità dell'esistenza di un quark dotato di incanto era stata prevista teoricamente quattro anni prima da Glashow, John Iliopoulos e Luciano Maiani. Tale quark è stato scoperto dal gruppo di Samuel C. C. Ting del Brookhaven National Laboratory e da quello di Burton D. Richter dello SLAC. La terza famiglia è entrata in scena nel 1975, con la scoperta della particella tau. Per lo più i fisici ritengono che il neutrino della tau sia distinto dal neutrino dell'elettrone, anche se ciò non è stato provato con certezza. Entro il prossimo quinquennio si prevede poi la scoperta del quark «alto», a meno che non si tratti di un'entità dotata di massa estremamente grande.

Non esiste alcun ostacolo teorico alla scoperta di nuove famiglie di particelle, e ciò potrebbe venire spiegato, secondo alcuni teorici, dalle teorie unificate delle interazioni deboli ed elettromagnetiche. Una di queste teorie è stata proposta alla fine degli anni sessanta indipendentemente da Steven Weinberg di Harvard e da Abdus Salam del Centro internazionale di fisica teorica di Trieste. Tale teoria interpreta esattamente molte delle proprietà delle interazioni deboli; nella sua formulazione attuale essa è totalmente compatibile con lo schema di classificazione in famiglie, ma non fornisce alcuna spiegazione dell'esistenza delle famiglie stesse. Altre formulazioni più estese, incorporanti la teoria di Weinberg e di Salam, danno questa spiegazione ma non hanno fondamento sperimentale. Inoltre, non esiste alcuna teoria convincente in grado di spiegare perché le masse delle particelle nelle varie famiglie abbiano proprio i valori riscontrati. La possibile proliferazione delle famiglie di particelle sta quindi per aprire un nuovo capitolo della fisica nucleare, mentre il lungo capitolo della ricerca dei costituenti fondamentali della materia potrebbe chiudersi con la definizione di un modello probante del confinamento dei quark e con la conferma della natura elementare dei quark e dei leptoni.

Una miniera di selce del Neolitico

L'avvento dell'agricoltura nell'Europa preistorica fece aumentare la richiesta di utensili di selce. In Olanda è stato scoperto un sito che, 5000 anni fa, produceva tonnellate di questo prezioso materiale

di Peter W. Bosch

La selce è stata per almeno mezzo milione di anni un materiale di importanza capitale nella vita del genere umano. Pur essendo più dura dell'acciaio, è anche fragile, cosicché si presta con facilità a essere ridotta in qualsiasi forma voluta mediante la percussione e la scheggiatura. I noduli di selce, alcuni dei quali misurano fino a un metro di diametro, sono particolarmente abbondanti negli strati calcarei che si depositarono sui fondali oceanici in tutto il mondo durante il Cretaceo superiore, circa 80 milioni di anni fa. In Europa strati calcarei, spinti verso l'alto dai movimenti tellurici e rimasti esposti in superficie, sono visibili in diverse località, dalla Francia e dall'Inghilterra meridionale fino alla Polonia e alla Cecoslovacchia più a est.

Durante il Paleolitico i cacciatori europei impiegarono in modo preferenziale come materia prima, per la lavorazione di asce, raschiatoi e altri utensili, i noduli di selce affioranti spontaneamente sulla superficie degli strati calcarei. Soltanto alcune migliaia di anni fa, con l'avvento dell'agricoltura come nuovo modo di vivere, la crescente esigenza di utensili di selce indusse gli europei del Neolitico a estrarre i noduli dal terreno invece di raccoglierli semplicemente sulla sua superficie. In questo articolo descriveremo quanto si è recentemente appreso riguardo a un sito minerario di questo genere, dal quale, circa 5000 anni fa, gli artigiani del Neolitico trassero il materiale per fabbricare oltre 150 milioni di utensili. Il sito, che si trova in Olanda nei pressi di Rijckholt, fu scoperto nel 1881 e da allora fu oggetto di studi a più riprese per tre quarti di secolo e infine sottoposto a meticolosi lavori di scavo da parte di archeologi dilettanti che diedero la loro opera per una notte alla settimana durante un periodo di nove anni.

Come fu scoperto questo importante sito? La Mosa, che nasce in Francia, nel suo percorso volto a settentrione verso il Mare del Nord attraversa il Belgio e poi bagna l'angolo più meridionale dell'Olanda, passando nei pressi di Maastricht. Nel 1881, durante un viaggio per ferrovia verso quella località dalla città belga di frontiera di Visé, Marcel de Puydt, un insigne archeologo belga, osservò che quella zona della valle della Mosa presentava delle somiglianze con altre località in cui, nel suo paese, aveva individuato siti preistorici. Il pendio boscoso saliva gradatamente dal fondo valle per trasformarsi infine in un altopiano ricoperto di foreste. De Puydt decise di esplorare la zona; e nello stesso anno, nei pressi di Rijckholt, si imbatté in un deposito superficiale di manufatti di selce.

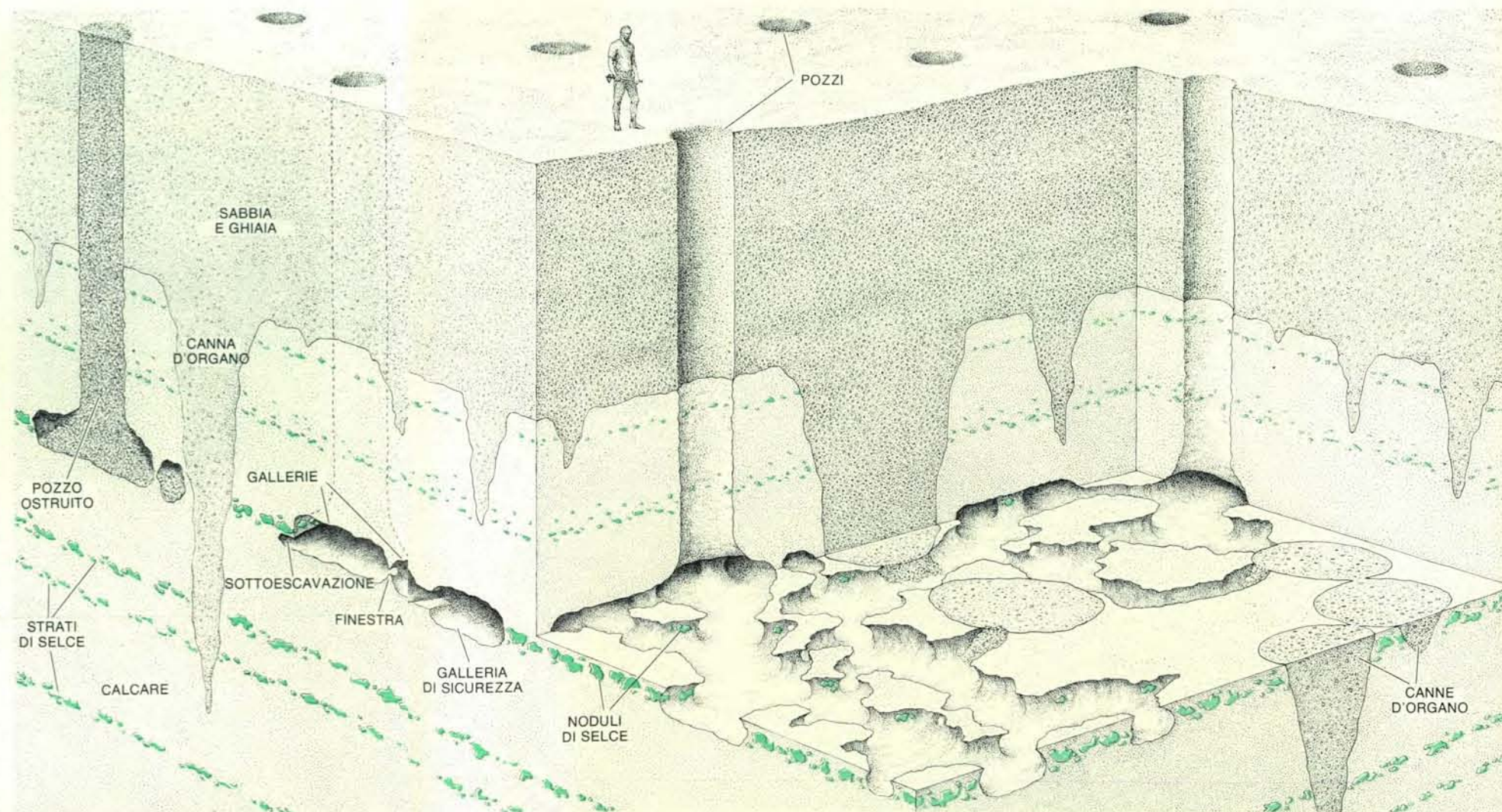
De Puydt è quindi da considerare lo scopritore del sito di Rijckholt, anche se soltanto in un secondo tempo egli ne trovò l'area di lavoro più importante situata in superficie. Si trattava di un bacino naturale, pieno di schegge di selce di scarto fino a una profondità di un metro e mezzo, al quale fu in seguito dato il nome di «Grand Atelier». Comunque, anche quando altri ricercatori vennero a esplorare il sito, de Puydt continuò a lavorare a Rijckholt a diversi intervalli, per un periodo di 35 anni, ampliando la sua raccolta di utensili di selce, tra i quali teste d'ascia, raschiatoi, punte per armi da lancio, lame di coltelli e utensili da miniera come i picconi ricavati dalle corna di cervo.

Anche il proprietario del terreno sul quale erano situate le miniere di Rijckholt, René de Geloes, effettuò delle ricerche in loco. Il conte de Geloes iniziò i lavori nel 1886; tra coloro che si associarono a lui nell'impresa vi era un giovane medico, Eugène Dubois, che quattro anni dopo scopriva a Giava il *Pithecanthropus erectus* (ora denominato *Homo erectus*). A Rijckholt lavorarono anche altri archeologi provenienti dalla Francia e dal Belgio; tra questi vi era Joseph Hamal Nandrin dell'Università di Liegi, i cui scavi iniziarono nel 1903 con l'aiuto degli abitanti del villaggio. Nel 1910 Nandrin si imbatté in parecchi pozzi minerari di antica data ostruiti da pietrisco calcareo. La scoperta di Nandrin fornì la prima sicura testimonianza che in Olanda durante la preistoria erano esistite miniere di selce (anche se il Grand Atelier di Rijckholt, contenente forti quantitativi di materiale siliceo di scarto, costituiva già una valida prova che i noduli di selce erano stati estratti nelle vicinanze).

Si dovette poi giungere fino al 1923 prima che un ricercatore olandese mostrasse interesse per le miniere di Rijckholt. Quell'anno A. E. van Giffen dell'Università di Groninga diede inizio a uno studio sistematico del sito. Egli fu il primo a effettuare una serie di scavi in trincea nel Grand Atelier ed ebbe cura di rilevare la posizione di tutti i suoi ritrovamenti. Quel metodo di scavo valse ben presto a dimostrare che gli artigiani del Neolitico avevano suddiviso il bacino in un certo numero di aree separate. Per esempio, le teste d'ascia erano state fabbricate in alcune zone, i coltelli in altre, i raschiatoi in altre ancora; ogni zona specializzata in una determinata lavorazione della selce poté essere identificata per la presenza di schegge di scarto caratteristiche di un particolare processo produttivo. Anche van Giffen esplorò il sito di Rijckholt in generale. Dove il margine del terreno pianeggiante scendeva fino a incontrare l'orlo del Grand Atelier egli trovò le tracce dei pozzi minerari. Sull'area pianeggiante stessa egli rinvenne sia cocci di vasellame sia resti di focolari e di perimetri di capanne di abitazione temporanea; tutto ciò lasciava supporre che i minatori si fossero accampati in quel sito durante gli scavi per l'estrazione della selce.

Nel 1926 van Giffen portò a termine la sua ricerca sul campo, ma, seguendo il suggerimento di alcuni archeologi belgi, i monaci del monastero domenicano di Rijckholt continuarono l'esplorazione del sito durante il loro tempo libero. Tra il 1929 e il 1932 anch'essi trovarono tracce di pozzi minerari sul terreno in pendio, nei pressi del Grand Atelier. Si imbatterono inoltre in gallerie minerarie (i tunnel orizzontali che vengono scavati verso l'esterno dal fondo di un pozzo) il cui soffitto era crollato, e ne trassero una quantià di utensili da miniera: oltre 100 picconi di corno di cervo e circa 1200 teste d'ascia di selce. Durante questo periodo, da Liegi giungeva a intervalli Nandrin per raccogliere manufatti per la sua università; egli continuò questa sua attività fino al 1953.

Nel decennio successivo a Rijckholt non ebbero luogo altri lavori. Nel 1964 un ricercatore olandese, H. T. Waterbolk dell'Università di Groninga, si recò nel sito come capo di un gruppo proveniente dall'Istituto di biologia e archeologia dell'università stessa. Il gruppo di Waterbolk operò sul terreno pianeggiante nel quale van Giffen aveva segnalato la presenza di profili di capanne. Con una ruspa venne asportato uno strato di terra dello spessore di un metro, su un'area di circa 4000 metri quadrati. La rimozione di questo strato superficiale, sovrappostosi nel tempo, rivelò che i perimetri delle capanne di van Giffen erano in realtà il risultato di fenomeni di subsidenza del suolo che ricopriva un gran numero di pozzi minerari del tipo scoperto per la prima volta da Nandrin nel 1910. Il numero dei pozzi (ne furono scoperti 66 nello spazio di 3000 metri quadrati) dimostrò che le miniere di Rijckholt avevano occupato un'area assai più vasta di quanto avessero supposto tutti i precedenti ricercatori.

Waterbolk e i suoi collaboratori, come van Giffen e i monaci domenicani prima di loro, trovarono le prove dell'esistenza

Questo spaccato di una parte della zona di estrazione della selce a Rijckholt, presso Maastricht, mostra a sinistra la sovrapposizione di uno strato di sabbia e ghiaia su un profondo deposito di calcare del Cretaceo, il quale contiene alcuni filoni di noduli di selce *(in colore)*. Le «canne d'organo» sono aree in cui l'acqua contenuta nel terreno ha disciolto il calcare provocando l'intrusione dello strato sovrapposto. La sezione verticale mostra un pozzo del Neolitico ostruito e le gallerie comunicanti con altri due pozzi. La sezione orizzontale, a destra, mostra due pozzi e la metà inferiore delle gallerie a essi collegate, secondo la mappa tracciata dai dilettanti che riaprirono le miniere nel 1964. La scala è fornita dalla figura umana.

Le principali miniere di selce del Neolitico comprendono Grimes Graves e Cissbury in Inghilterra, Grand Pressigny e Saint-Mihiel in Francia, Spiennes in Belgio, Hov in Danimarca, Krzemionki in Polonia e Sumeg in Cecoslovacchia. Il rettangolo delimita l'area che appare ingrandita in basso.

Le miniere di selce olandesi sono situate fra Maastricht e Aquisgrana, pressoché al centro di un esteso deposito di calcare *(in colore)*. Fu l'archeologo belga Marcel de Puydt nel 1881 ad accertare che in quest'area gli uomini preistorici fabbricarono utensili di selce in grande quantità.

di gallerie minerarie crollate lungo il pendio attiguo al Grand Atelier. Essi tentarono di scavare un passaggio nelle gallerie neolitiche, ma non vi riuscirono, in quanto nessuno di essi aveva la minima esperienza nel campo minerario. La ricerca si arenò; sembrava che ogni interesse a scoprire la reale entità delle miniere di Rijckholt fosse caduto.

A questo punto un geologo di professione, che aveva come dilettante un grande interesse per l'archeologia, avvicinò Waterbolk con una nuova proposta. Il nuovo venuto, Werner M. Felder, era un membro attivo del capitolo di Limburg della Società geologica olandese. Egli propose di scavare un tunnel nel pendio sovrastante il Grand Atelier, con lo scopo di seguire il filone calcareo ricco di selce che i minatori del Neolitico avevano sfruttato praticando dei pozzi nel pianoro soprastante. Il tunnel avrebbe certamente intersecato molte delle gallerie orizzontali che i minatori avevano scavato nello strato calcareo per estrarne i noduli di selce. Secondo i calcoli di Felder sarebbe stato sufficiente che il tunnel esplorativo raggiungesse la lunghezza di 140 metri. Egli suggerì inoltre l'opportunità di esplorare anche i primi 10 metri di ogni galleria preistorica che il tunnel avrebbe tagliato.

Waterbolk approvò il piano di Felder e offerse la sua partecipazione come supervisore scientifico. Le miniere di Rijckholt ora non appartengono più agli eredi del conte de Geloes, ma fanno parte di una riserva naturale locale; il responsabile della riserva ha dato la sua approvazione a uno scavo sperimentale e 13 membri del capitolo di Limburg della Società geologica formarono di lì a poco un ente a cui diedero il nome di Gruppo di ricerca delle miniere preistoriche di selce. Il gruppo comprendeva geologi, lavoratori, studenti, un insegnante e, fatto non meno importante, alcuni esperti minatori provenienti dalle vicine miniere di carbone. Nacque così la più piccola e la meno ufficiale compagnia mineraria olandese.

Il 16 giugno 1964, quando demmo inizio ai lavori, nessuno poteva prevedere l'importanza che l'impresa poteva assumere, quanto sarebbe durata e quale ne sarebbe stato il costo. A tale proposito aggiungerò che nessuno immaginava quale aspetto avrebbero avuto le miniere neolitiche quando le avremmo raggiunte. L'opinione generale era che le gallerie preistoriche sarebbero state abbastanza alte da consentirci di camminare eretti: in tale ottimistica previsione, alcuni membri vennero anche equipaggiati di tutto il necessario per tracciare delle mappe. Ma le cose si svolsero diversamente.

La nostra prima iniziativa fu quella di costruire una baracca nel punto di ingresso sul pendio, per avere un riparo dalle intemperie. Fu una saggia precauzione, perché il tempo era cattivo quando, con pale e picconi, attaccammo gli scavi effettivi per aprire l'imboccatura di un tunnel su due metri di larghezza per un metro e mezzo di altezza. Come volontari non retribuiti, potevamo dedicare agli scavi sol-

tanto i fine settimana. Appena il tunnel si inoltrò di qualche metro, comunque, avevamo già incontrato le prime gallerie neolitiche e, dopo alcune settimane di attività, avevamo raggiunto il primo pozzo.

Tanto le gallerie quanto il pozzo erano ostruite completamente dal pietrisco calcareo; inoltre le gallerie erano parzialmente crollate. Dopo la rimozione del pietrisco introducemmo delle travi per puntellare le gallerie e i punti di accesso al pozzo. Mentre eravamo intenti allo sgombero delle gallerie trovammo le prime teste d'ascia di selce. Durante la rimozione della colmata di pietrisco dal primo pozzo fu rinvenuta pure una discreta quantità di carbone, verosimilmente il residuo di legna caduta da un focolare preistorico situato presso la sommità del pozzo; questo reperto fu inviato in laboratorio per l'analisi del carbone 14 al fine di determinare l'età.

Galleria di una miniera, alta appena 60 centimetri, sgomberata in parte dai detriti calcarei ammassati dai minatori neolitici. Le inclusioni scure, visibili nella parete sinistra, sono noduli di selce.

Due «finestre» aperte nelle pareti di un'altra galleria. Di dimensioni troppo ristrette per consentire il passaggio da una galleria all'altra, sono probabilmente formazioni di carattere accidentale.

Attendemmo ansiosamente il responso. Il sito di Rijckholt era stato sottoposto a esplorazioni per 80 anni, ma nessuno conosceva con esattezza l'età della miniera. La lettura che ci pervenne forniva la data 3159 (± 60 anni) a. C. Evidentemente l'estrazione della selce aveva avuto inizio nell'Olanda meridionale poco più di 5000 anni fa.

Lo scavo esplorativo aveva fatto inoltrare il nostro tunnel per 15 metri nello strato calcareo. I risultati erano abbastanza incoraggianti da giustificare un progetto su vasta scala, anche se ancora a livello dilettantistico, cosicché fu necessario richiedere un permesso ufficiale del governo. Esso fu ottenuto in breve tempo, subordinato però all'impiego di supporti in acciaio in luogo di quelli in legno. La Fondazione olandese per la ricerca scientifica poi erogò un sussidio che ci fu di aiuto per affrontare le spese per il costo più elevato del materiale da usare.

Ora che il progetto aveva veste ufficiale, occorreva stabilire alcune regole formali. I gruppi di lavoro addetti allo scavo esplorativo si erano sinora recati sul posto di sabato, ma la maggior parte di noi era costituita da uomini sposati, che desiderava passare il fine settimana con la famiglia. Fu deciso quindi di prolungare la nostra settimana lavorativa con le ore del venerdì notte, e precisamente dalle 19 del venerdì alle 3 del sabato. Fu anche concordato che ogni membro del gruppo avrebbe lavorato per un minimo di 20 notti all'anno. Infine 20 volontari, che aderirono a tale programma, formarono il quadro di base del gruppo minerario del capitolo di Limburg, cosicché l'esecuzione del lavoro notturno a Rijckholt poté essere affidata a due squadre distinte di 10 uomini ciascuna, per 40 delle 52 settimane dell'anno. Anche quando il venerdì cadde nella vigilia di Natale o di Capodanno, la maggior parte dei volontari si presentò al lavoro.

Werner Felder e suo fratello Joseph furono scelti rispettivamente come capo e direttore tecnico del progetto. Ambedue avevano lavorato nelle miniere di carbone e avevano quindi esperienza pratica in materia. Altri membri del gruppo di lavoro si assunsero il compito di tenere i contatti con le varie istituzioni e musei che avevano preso interesse al progetto.

In breve, ecco la sequenza dei nostri progressi. Nel 1965 furono portate a mano 50 tonnellate di acciaio per sostituire le travi di supporto impiegate durante lo scavo sperimentale. Vennero collocate delle rotaie per poter trasportare il materiale calcareo di scarto fuori della miniera su carrelli anziché con le carriole a mano. Verso la fine dell'anno il tunnel aveva raggiunto la lunghezza di 18 metri.

Nel 1966 una delle compagnie che si interessavano al progetto ci inviò un

Il tunnel esplorativo, scavato da volontari durante i fine settimana tra il giugno del 1964 e il dicembre del 1972 è qui visibile in sezione longitudinale e trasversale *(all'estrema destra)*. La scala è fornita dalla figura umana. Con una estensione di 150 metri dall'imboccatura, situata nei pressi del Grand Atelier, il percorso del tunnel seguiva lo strato calcareo di 30 centimetri di spessore, che i minatori avevano trovato particolarmente ricco di noduli di selce. Gli scavatori raggiunsero le prime gallerie neolitiche dopo pochi metri; alla fine di quei nove anni essi avevano individuato 66 pozzi minerari, da ognuno dei quali si diramavano gallerie la cui superficie era mediamente di 25 metri quadrati. Si ritiene che l'intera area mineraria di Rijckholt abbia fornito 41 250 metri cubi di selce, sufficienti a fabbricare 153,3 milioni di teste d'ascia.

compressore ad aria e dei martelli pneumatici, che ci avrebbero consentito di aprire un varco nello strato calcareo, sempre più duro, con maggior facilità. Durante l'anno precedente la nostra unica fonte di illuminazione elettrica era stata costituita da una batteria da 12 volt; ora eravamo in grado di avere in prestito un piccolo generatore e di predisporre di un sistema di illuminazione di maggior ampiezza. Col progredire dei lavori divenne necessario suddividere i nostri gruppi del venerdì notte in squadre più ridotte: seguendo il metodo in uso nelle miniere carbonifere, ogni squadra era diretta da un capo, che era responsabile della sicurezza e aveva inoltre il compito di registrare tutti i ritrovamenti. Quell'anno il tunnel raggiunse i 38 metri di lunghezza.

Nel 1967 il trasporto del materiale calcareo si fece sempre più difficoltoso. L'industria mineraria olandese che aveva mostrato grande interesse al nostro progetto ci fece pervenire un nastro trasportatore lungo 150 metri. Poiché al sito non giungeva alcuna linea elettrica, ci procurammo un trattore e ne collegammo con una cinghia una ruota posteriore all'albero motore del trasportatore. Grazie ai sussidi in denaro e agli equipaggiamenti avuti in dono, ora ciascuno di noi era dotato di un idoneo completo da minatore e di un elmetto di sicurezza fornito di lampada. Ogni membro del gruppo fu inoltre assicurato contro gli incidenti, giacché il lavoro non era privo di rischi. Sul finire di quell'anno il tunnel era lungo 67 metri.

Nel 1968 si continuò a lavorare a squadre. Una di esse fu impiegata nell'opera di prolungamento delle gallerie. Una seconda fu addetta ai rifornimenti, alla manutenzione delle attrezzature e allo sgombero del sito. Una terza squadra, e talvolta una quarta, furono incaricate dello sgombero del pietrisco dalle gallerie laterali, di tracciare le mappe delle nuove aree e di registrare i manufatti e il carbone di legna che venivano scoperti. Alla fine dell'anno il tunnel raggiunse i 108 metri.

Nel 1969, sentito anche il parere di Waterbolk, il gruppo decise di prolungare il tunnel per l'intera estensione dell'area mineraria neolitica. Tale decisione richiedeva l'ulteriore scavo di altri 40 metri e tutti i membri del gruppo si mostrarono entusiasti di questa prospettiva. Poiché le squadre operative si concentrarono tutte sullo scavo del tunnel, fu tralasciato l'esame di alcune gallerie appena scoperte. Tuttavia prima che l'anno volgesse al termine, circa tre quarti di tutte le gallerie della miniera erano state sgombrate dal pietrisco per una lunghezza di 10 metri. A questo punto erano stati recuperati circa 10 000 utensili di selce. Alla fine dell'anno il tunnel era di circa 110 metri.

Nel 1970, 1971 e 1972 i lavori proseguirono sia agli scavi sia a tavolino. Quest'ultima attività comprendeva la raccolta di fondi (a quell'epoca il governo aveva incominciato a chiudere molte delle miniere di carbone che costituivano un sostegno per la nostra impresa, facendo così venir meno la loro assistenza), la corrispondenza, la stesura delle mappe e la compilazione di un catalogo dei reperti. I pochi metri di tunnel che ci restavano ancora da scavare ci posero di fronte a un vecchio problema su nuova scala.

«Canna d'organo» è il nome attribuito a un pozzo formato da infiltrazioni d'acqua presente nel terreno attraverso lo strato calcareo. Le canne d'organo hanno un diametro variabile da uno a oltre 10 metri; lo spazio in precedenza occupato dal calcare risulta riempito da sabbia e ghiaia. Se durante lo scavo ci si imbatte in una canna d'organo, vi è un serio pericolo che quel materiale incoerente si riversi nel tunnel bloccando il passaggio e, in qualche caso, ostruendolo completamente. In tutta l'area di Rijckholt le canne d'organo erano presenti, ma fortunatamente ne incontrammo soltanto qualcuna di piccole dimensioni. Ora che eravamo quasi giunti alla conclusione del nostro lavoro, dovemmo affrontarne alcune pericolosamente più grandi. Tuttavia, prima della fine del 1972 il tunnel aveva raggiunto la lunghezza prestabilita di 150 metri.

Nel 1973, avendo ormai portato a termine le operazioni di scavo, fummo in grado di dedicare un tempo maggiore allo sgombero delle restanti gallerie. Anche queste attività furono completate entro la fine dell'anno. Joseph Felder calcolò che avevamo aperto un'area mineraria di 3000 metri quadrati. Tutto ciò che ci restava da fare era di portar via il macchinario e gli equipaggiamenti e di riportare la superficie del sito alle condizioni originarie. Durante tutti quegli anni il personale della riserva naturale ci era stato di aiuto in quest'opera di restauro, trasportando altrove i residui del calcare man mano che veniva da noi rimosso dalle gallerie della miniera. Dopo aver dedicato per 9 anni i nostri weekend a questa fatica sotterranea, potevamo ora rivolgere la nostra attenzione all'elaborazione dei dati finali.

Che cosa rappresentavano in epoca neolitica le miniere di Rijckholt? Dal ritrovamento di due altri campioni di carbone di legna sappiamo che l'estrazione della selce nel sito era già in corso più di 5000 anni fa. (Le loro datazioni, 3120 e 3150 a.C., concordano in modo soddisfacente con quella eseguita con il carbonio 14 sul nostro primo reperto.) Non siamo in grado di stabilire la durata nel tempo dell'attività mineraria, né l'entità della mano d'opera che vi fu impiegata. Tuttavia Joseph Felder ha potuto formulare alcuni dati statistici generali in proposito. Anzitutto, l'area mineraria di Rijckholt aveva un'estensione di 25 ettari. Durante l'esplorazione di soli 3000 metri quadrati di tale area ci siamo imbattuti in 66 pozzi minerari. Se tale campione di territorio avesse valore rappresentativo, se ne potrebbe dedurre che l'intera area di Rijckholt doveva contenerne circa 5000. (Si considera come unità mineraria ogni pozzo con le gallerie a esso collegate.) Secondo la stima di Felder, le 5000 unità di Rijckholt fornirono complessivamente 41 250 metri cubi di noduli di selce, cioè la materia prima sufficiente a fabbricare 153,3 milioni di teste d'ascia. Se prendiamo in considerazione l'ipotesi che l'attività estrattiva si sia protratta per 500 anni, la quantità media di produzione sarebbe stata di 1500 teste d'ascia al giorno, supponendo che tanto i minatori quanto i fabbricanti di utensili di selce del Grand Atelier lavorassero durante l'intero anno e non soltanto stagionalmente.

I pozzi che penetravano nel terreno pianeggiante, scendendo fino allo strato calcareo che conteneva la selce, avevano profondità variabili da 10 a 16 metri, a seconda dello spessore della sabbia, della ghiaia e del loess che si erano depositati al di sopra del calcare. Il diametro dei pozzi variava da un metro a un metro e mezzo. Mentre procedevano con gli scavi attraverso il materiale incoerente dello strato superiore, i minatori, allo scopo di prevenire eventuali crolli, costruivano pareti di sostegno con rami d'albero intrecciati: per tutta la lunghezza della parete di ogni pozzo da noi sgomberato, sino alla superficie, erano infatti visibili dei solchi simili a impronte di rami. (Noi stessi ci mettemmo al riparo dalla possibilità di crolli collocando in successione nei pozzi, via via che gli scavi proseguivano, delle armature di acciaio.)

In media le gallerie aperte dai minatori di Rijckholt alla base di ogni pozzo avevano una superficie complessiva di 25 metri quadrati. L'altezza delle gallerie era di circa 60 centimetri, mentre l'ampiezza variava a seconda dell'ubicazione dei noduli di selce. Il metodo di estrazione era identico a quello odierno noto come sistema «camera e pilastro». Il filone calcareo contenente la selce, dello spessore di 30 centimetri circa, veniva scavato dal di sotto e poi rimosso, un blocco alla volta; a intervalli il calcare era lasciato intatto, in modo da formare dei pilastri allo scopo di sostenere la volta della galleria. Per integrare la funzione di questi pilastri naturali i minatori riempivano nuovamente l'area scavata in precedenza con il materiale calcareo di scarto. Quando una galleria era stata completamente sfruttata, veniva usata come luogo di scarico per i detriti calcarei estratti da quella successiva: il 95 per cento delle gallerie da noi scoperte risultavano essere state riempite in un secondo tempo in base a questo criterio.

Come riuscivano i minatori ad aprirsi un varco attraverso una quantità incalcolabile di tonnellate di calcare? Gli scavi erano effettuati con le asce a martello fabbricate dai loro compagni che lavoravano la selce in superficie, nel Grand Atelier. A giudicare da certi spazi cavi da noi osservati nei cumuli di detriti calcarei, le teste d'ascia erano di tipo notevolmente uniforme; misuravano tra i 15 e i 18 centimetri di lunghezza per 5 centimetri di larghezza e 3,5 di spessore.

Scavare lo strato calcareo doveva essere impresa assai dura e richiedeva una considerevole abilità; riteniamo che un minatore logorasse almeno cinque teste d'ascia per riuscire a rimuovere un metro cubo di calcare. Quando le asce risultavano smussate, venivano affilate sul posto, impiegando percussori di pietra calcarea per ritoccarne il lato tagliente. A tale scopo veniva usato un tipo di calcare ricristallizzato, particolarmente duro. Trovammo la maggior parte dei percussori litici insieme a gruppi di 10-20 teste d'ascia, e a queste associazioni demmo il nome di depositi di asce. I siti in cui si provvedeva all'affilatura delle lame smussate si trovavano generalmente nelle vicinanze di un pozzo ed erano caratterizzati dalla presenza di una quantità di schegge di scarto.

Nelle miniere trovammo oltre 15 000 manufatti, la maggior parte dei quali era costituita da asce smussate o spezzate. In base alle estrapolazioni di Joseph Felder ciò indicherebbe che nelle gallerie ancora inesplorate si troverebbero altri due milioni e mezzo di asce, abbandonate dai minatori di Rijckholt. Se ciò fosse esatto, significherebbe che essi reinvestirono meno del 2 per cento della loro produzione complessiva di selce nell'attività estrattiva, poiché il quantitativo richiesto per lo sfruttamento di ogni miniera si aggirava intorno alle 350 asce. Furono da noi rinvenuti anche alcuni picconi di corno di cervo, simili a quelli che de Puydt aveva trovato a Rijckholt in superficie durante le ricerche da lui effettuate durante il diciannovesimo secolo. Il fatto che ne abbiamo rinvenuti in numero così esiguo lascia supporre che nello scavo delle gallerie se ne facesse scarso uso, perché a quel livello il calcare risultava più duro delle corna di cervo.

Sembra probabile che i minatori di Rijckholt lavorassero contemporaneamente in più di una miniera. Dà valore a tale ipotesi la nostra scoperta dell'esistenza di «gallerie di sicurezza». Questi passaggi non erano ostruiti, o lo erano soltanto in modo parziale, dai detriti calcarei: essi servivano a porre in comunicazione fra loro due e talvolta più pozzi. Questo sistema di collegamenti non dovette richiedere molto lavoro supplementare da parte dei minatori, in quanto i pozzi stessi erano stati scavati a breve distanza l'uno dall'altro, probabilmente allo scopo di

limitare lo spazio da percorrere per trasportare i noduli di selce dalle gallerie ai pozzi. Risulta in modo chiaro che l'ultima galleria di ogni unità mineraria si prolungava fino a raggiungere il pozzo successivo. L'uso regolare delle gallerie di sicurezza è reso evidente dalla compattezza del pietrisco calcareo sul fondo e dalle scalfitture e graffiature visibili sulle pareti e sulle volte, prodotte dal passaggio frequente dei minatori che vi procedevano carponi.

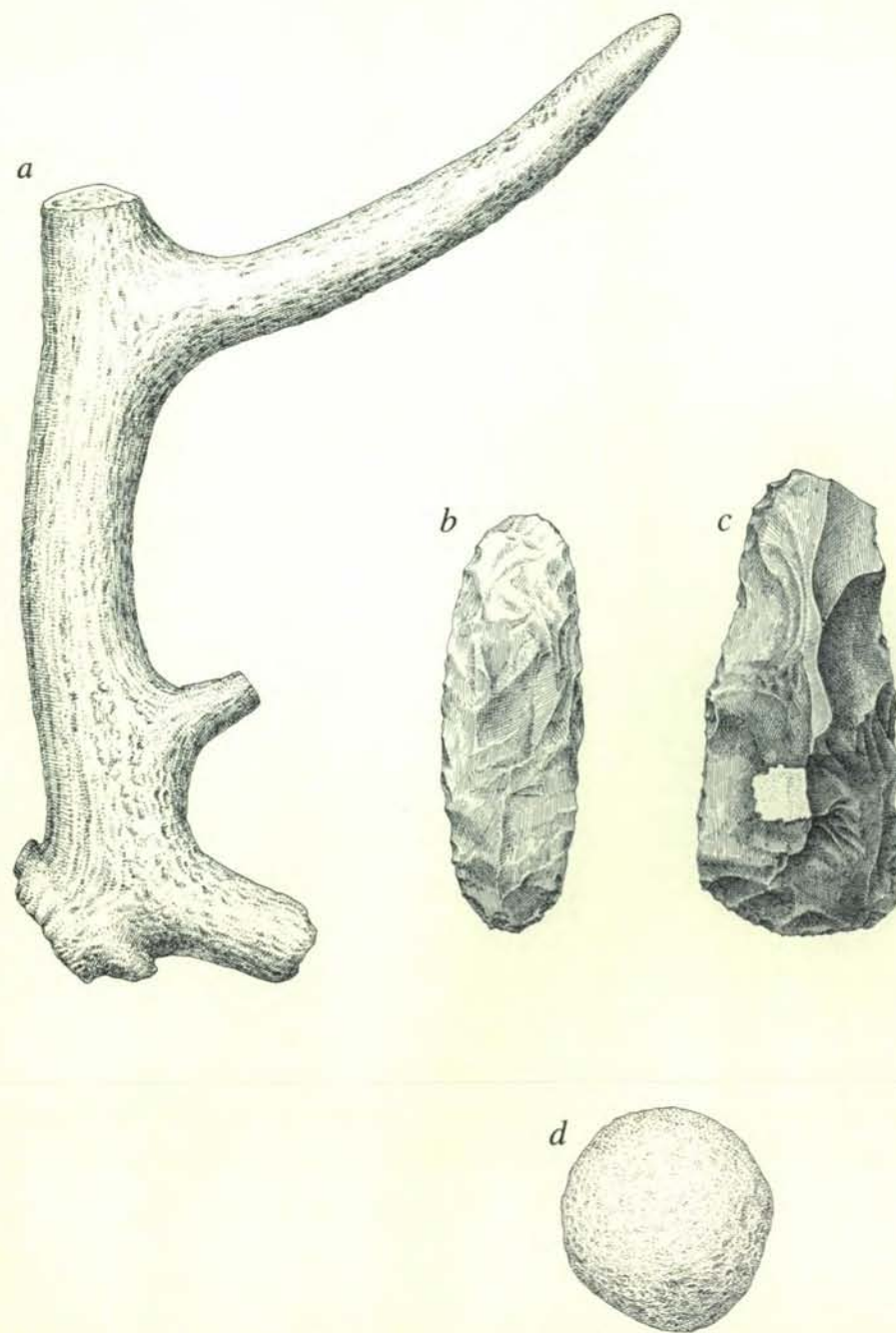
Non siamo in grado di stabilire con certezza in qual modo gli uomini scendessero al fondo dei pozzi e potessero poi risalirne, e neppure quale fosse il metodo usato per portare in superficie i noduli di selce. Sembra logico supporre in ambedue i casi che facessero ricorso alle funi, ma l'unica prova al riguardo è data dalla presenza di profondi solchi scavati nel calcare all'imboccatura dei pozzi e delle gallerie.

Non abbiamo trovato tracce relative alla presenza di lanterne o torce nel sottosuolo; le loro fiamme avrebbero probabilmente sottratto ai minatori quell'ossigeno che era loro più necessario dell'illuminazione. I depositi di asce erano situati presso il fondo dei pozzi, forse perché i minatori avevano bisogno di luce per affilare le teste d'ascia. Lavorando in galleria, comunque, i minatori dovevano evidentemente accontentarsi di quel poco di illuminazione indiretta che riusciva a filtrare fino a loro dall'alto.

Durante il primo anno dei nostri scavi in galleria su grande scala ci imbattemmo in una testimonianza alquanto enigmatica che doveva essere in relazione con qualche culto praticato dai minatori. Il 5 novembre 1965, al termine di una galleria appena sgomberata, trovammo una fossa poco profonda che conteneva un cranio umano privo della mascella inferiore. Eccettuata l'area in cui era stato rinvenuto il cranio, la galleria era stata riempita completamente, dopo l'uso, con il pietrisco calcareo. Era evidente che soltanto il cranio era stato deposto nella fossa, probabilmente come atto rituale di una sepoltura secondaria. Il reperto fu esaminato da G. van Vark, del Laboratorio di anatomia ed embriologia dell'Università di Groninga, che lo ha identificato per quello di un individuo di sesso maschile di circa 40 anni; esso presenta lesioni intorno alla tempia e all'occhio sinistri, ma il fatto che l'osso si fosse rigenerato indica che non furono le lesioni la causa della morte. A quale pratica esattamente sia da attribuire la presenza del cranio nella miniera non è facile immaginare. Si può soltanto ipotizzare che sia stata qualche forma di venerazione ancestrale a fornirne il motivo.

Nel 1975 si tenne a Maastricht il secondo di una serie di simposi internazionali sul tema della selce e una visita a Rijckholt faceva parte del programma. A quell'epoca gli scavi erano stati abbandonati da tre anni, i supporti in acciaio del tunnel si stavano coprendo di ruggine e anche le gallerie mostravano segni di deterioramento. I partecipanti al simposio promossero l'invio di un rapporto al ministro olandese della cultura, facendo rilevare che le miniere di Rijckholt erano un monumento preistorico della massima importanza, e sollecitando dal governo un'azione intesa alla sua conservazione. Oltre al governo, si fecero avanti molte imprese private con offerte di aiuto. A partire da quello stesso anno erano già state prese numerose iniziative per la preservazione dell'area già esplorata con gli scavi.

Oggi il pavimento del tunnel è stato scavato più in profondità, in modo che i visitatori possano camminarvi in posizione eretta per la sua intera lunghezza, mentre le pareti e la volta sono state protette dalle infiltrazioni d'acqua. È stata costruita una piccola tettoia per ricevere i visitatori all'ingresso del tunnel e la luce elettrica illumina tanto il tunnel che le gallerie. Turisti e studiosi possono entrarvi e rendersi conto delle fatiche affrontate dai loro antenati di 5000 anni fa per l'estrazione della selce, di cui avevano bisogno per fabbricare i loro utensili, ma anche del notevole risultato conseguito da 20 archeologi dilettanti che dedicarono a questa impresa soltanto le ore notturne di tanti venerdì.

a *b* *c* *d*

Gli utensili dei minatori rinvenuti nelle gallerie di Rijckholt comprendono alcuni picconi di corno di cervo *(a)*, teste d'ascia *(b, c)* e percussori per la riaffilatura delle teste d'ascia smussate *(d)*. Alcuni percussori erano fatti di quarzo e altri, come quello visibile nella figura, di un calcare particolarmente resistente. Si ritiene che la rimozione di ogni metro cubo di calcare ponesse fuori uso 5 teste d'ascia, il che equivale a un consumo di 350 teste d'ascia per ciascuna miniera interamente sfruttata.

GIOCHI MATEMATICI

di Martin Gardner

Come diventare un medium anche se siete un cavallo o un altro animale

La prestidigitazione è l'arte di intrattenere il pubblico facendo cose che apparentemente violano le leggi naturali, come far levitare una ragazza, volatilizzare un elefante, materializzare colombe o piegare cucchiai sfiorandoli con un dito, e così via.

Un piccolo settore della magia, cresciuto sull'onda del successo di sedicenti medium, non si interessa tanto alla produzione di fatti impossibili; quanto a quelli che in qualche modo sono estremamente improbabili. Imprese di questo genere vengono solitamente eseguite con quei mezzi che servono a introdurre il caso nei giochi d'azzardo: dadi, mazzi di carte e via discorrendo. In tal modo si crea una sovrapposizione tra i metodi usati dai medium ciarlatani e quelli di abili giocatori e ciarlatani onesti (maghi).

Un esauriente resoconto dei rapporti tra magia e probabilità richiederebbe un'enciclopedica disamina dei modi per controllare dadi, carte, monete lanciate in aria, ruote e palline delle roulette, e i vari altri elementi dei giochi d'azzardo. In tutti questi casi è possibile barare sfruttando stratagemmi incredibilmente banali.

Questo mese passerò velocemente in rassegna alcuni degli innumerevoli metodi con cui un falso medium può propagandare la propria abilità quando usi carte da gioco o carte ESP. Oggi tutti i parapsicologi seri vogliono essere sicuri che il medium non veda o maneggi le carte, ma agli albori della parapsicologia, quando si ottennero i più sensazionali risultati, molto spesso queste precauzioni non venivano rispettate. Anche oggi nelle prove non rigorose e nelle dimostrazioni pubbliche di medium disonesti le carte ESP sono talvolta in piena vista.

Come può un abile medium ottenere punteggi superiori a quelli casuali? Una tecnica è quella di marcare nascostamente le alcune carte mentre le si maneggia nel corso dei test preliminari. Ci sono molti modi di farlo. I giocatori ricorrono spesso a quello che in gergo si chiama «intonaco», una sostanza cerosa che può essere messa sul retro di un bottone della giacca in modo che sia facilmente raggiungibile con la punta di un dito. L'intonaco lascia sulla carta una macchia molto leggera indistinguibile da quelle che si formano durante le normali manipolazioni delle carte. Un altro metodo è quello di fare una piccola tacca su un angolo della carta con l'unghia di un dito. Un'altra possibilità è quella di grattare con l'unghia del pollice un angolo così da formare una linea bianchiccia. Basta marcare poche carte per ottenere un punteggio significativo.

Talvolta non è nemmeno necessario segnare le carte per riconoscerle dal dorso. Con l'uso sia le carte da gioco che le carte ESP acquistano tutta una serie di piccole imperfezioni, piccole macchie di sporco, angoli leggermente piegati, ecc. È questo il motivo per cui i giocatori professionisti cambiano continuamente i mazzi. Se un un giocatore di poker particolarmente osservatore nota una piccola macchia di sporco sull'angolo di una carta e in una mano successiva la vede su una carta di un avversario, è sveglio o bara?

Non è generalmente noto, neppure tra i maghi, che le carte ESP ufficiali attualmente in uso (autorizzate da J. B. Rhine) hanno il dorso «orientato». Vale a dire, esaminandole attentamente si può notare che il dorso non resta identico quando le si ruoti di 180°. Per esempio, l'angolo superiore destro di una carta ESP può avere o meno una stella. Nel corso di una mano di prova ci sono diversi modi in cui un medium può disporre un mazzo di 25 carte in modo che risultino poi con un'unica orientazione. Così, potrà cercare di indovinare le carte che sono orientate in un verso, tralasciando le altre; le carte in questione vengono collocate in un mucchio a parte. Dopo la prova (che avrà dato un punteggio di livello casuale) il mucchio viene rigirato prima di ricostituire il mazzo. Le carte avranno ora tutte l'orientamento voluto e il medium è pronto a fare miracoli.

La letteratura sulle carte magiche abbonda di trucchi basati sul principio dell'orientamento. Il medium può aprire a ventaglio il mazzo di carte coperte collocate sul tavolo e invitare qualcuno che sia seduto all'estremità sinistra del tavolo a estrarre una carta. Successivamente costui è pregato di passarla a qualcuno seduto all'estremità destra del tavolo, perché ne controlli il valore. La seconda persona rimette la carta nel mazzo e lo rimescola. Chiaramente la manovra capovolge la carta (ci sono molti altri sottili espedienti per girare le carte orientate). Le carte vengono poi disposte in una fila. Il medium gira attorno, fa passare lentamente la mano sotto la fila di carte per «sentire le vibrazioni» e identifica facilmente la carta scelta. Questa viene scoperta facendola ruotare lungo l'asse minore per venire poi rimessa nella fila coperta girandola di lato. In questo modo le carte sono di nuovo orientate.

Non è necessario disporre di un mazzo ESP uniformemente orientato per ottenere questo risultato. Supponiamo che i dorsi delle carte siano disposti a caso e che voi siate un medium. Disponete cinque carte in fila e memorizzate i dorsi come se fossero un numero binario, diciamo 11010. Seguite la strategia precedente per selezionare una carta e rimetterla col dorso girato. Potrete facilmente trovare la carta e ripetere l'esperimento quante volte vorrete.

Un mazzo standard di carte ESP è composto da cinque carte per ogni simbolo: una stella, una croce, un cerchio, delle linee ondulate e un quadrato. È sufficiente avere le dieci carte di due simboli orientate in un modo e le 15 carte per gli altri tre simboli nell'altro per ottenere punteggi significativi quando si cerchi di indovinare le carte di un mazzo mescolato distribuito in modo da esibire il dorso. Con un mazzo non preparato il punteggio atteso è di cinque successi, supponendo che non si abbiano delle informazioni sugli esiti fino a quando non sia completata la distribuzione. Con due simboli orientati avete una probabilità di 1/2 di indovinare correttamente fra i due. Ciò corrisponde a un'attesa di 10/2 ovvero 5 successi. Avete poi la probabilità di 1/3 di scegliere correttamente fra gli altri tre simboli, ossia un'attesa di 15/3, cioè cinque successi. L'attesa totale è dunque di cinque più cinque, ossia 10 successi per un giro delle 25 carte. Se il giro è ripetuto quattro volte, vi aspetterete per questo test standard 40 successi su 100. I parapsicologi considerano un punteggio di 30 come indicativo di un eccellente ESP, e un punteggio di 40 addirittura sensazionale. Questo è molto più impressionante di un punteggio pieno, che indurrebbe a sospettare fortemente un imbroglio. Si noti il fatto curioso che se viene marcato un solo simbolo l'aspettativa è già di quaranta successi su 100. Ciò tuttavia comporta che un simbolo verrà sempre individuato correttamente; le registrazioni sul foglio dei punteggi rivelerebbero perciò il trucco.

Se in una dimostrazione le carte non sono alla vista del medium, è possibile ricorrere a quell'imbroglio che i maghi chiamano «della spalla»: qualcuno nascosto dietro a uno schermo manda segnali convenzionali al medium con tecniche poco conosciute. (Una delle più difficili da individuare ricorre a un emettitore di impulsi elettronici che il medium riceve grazie a un apparecchio nascosto nell'anello. Pochi degli esaminatori di medium tengono conto di questa possibilità.)

L'uso di spalle è molto comune nella storia dei fenomeni medianici. Quasi tutti gli animali sapienti, tra i quali si ricordano cavalli ammaestrati, cani e anche maiali, erano guidati da segnali sonori segreti che possono essere facilmente percepiti da specie che hanno una sensibilità uditiva migliore di quella dell'uomo. Per esempio, l'istruttore ha scelto una carta da gioco e la indica con segnali magici convenzionali. L'animale «controlla» le 52 carte disposte in un ampio cerchio o su un lungo cavalletto e quando arriva alla carta giusta l'annusa o la prende in bocca. La carta è naturalmente identificata da un segnale che solo l'animale può sentire, e che può andare da un leggero sternuto allo schiocco di un'unghia contro un'altra di una mano tenuta in tasca o dietro la schiena. Non è detto che il segnale sia fatto dalla persona sulla scena, può anche essere qualcuno seduto in prima fila.

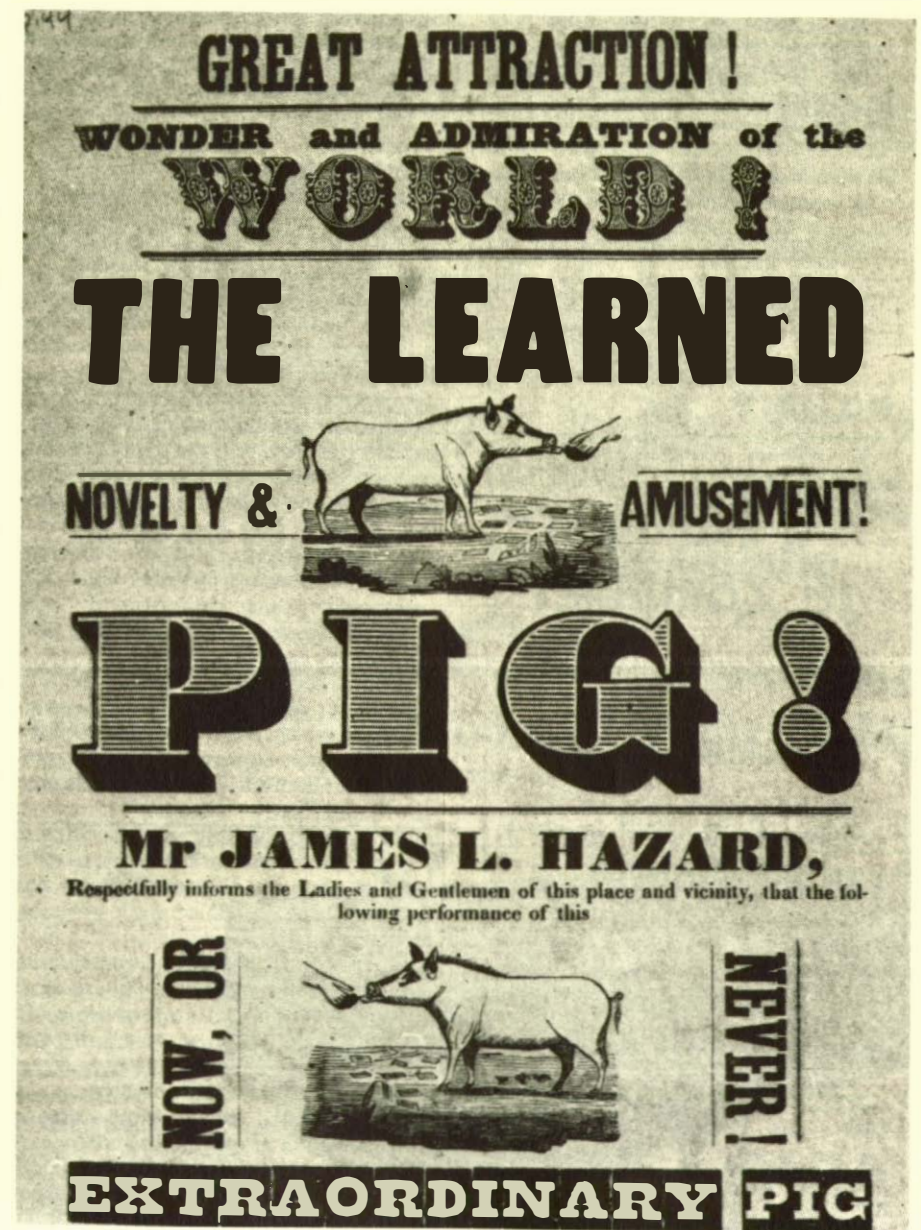
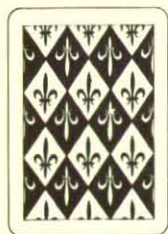
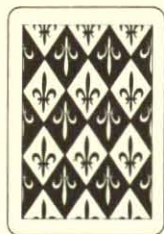
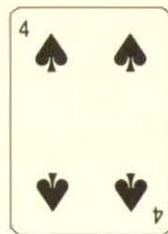
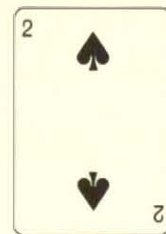
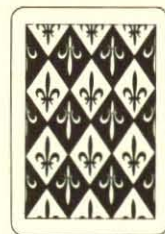
L'animale addestrato può così scegliere le carte che formano la risposta a una domanda o a un problema matematico posto da uno spettatore. Molti vecchi libri forniscono numerosi dettagli sul modo di addestrare gli animali a questo scopo. Uno dei migliori è *Haney's Art of Training Animals* (Jesse Haney & Co., Publisher, 1869), che raccomanda i segnali con le unghie. Un altro è *The Expositor: Or Many Mysteries Unravelled* di William Frederick Pinchebeck (pubblicazione privata, 1805). Contiene un eccellente capitolo dedicato all'addestramento di maiali sapienti con rumori nasali.

Ai tempi della sua giovinezza, J. B. Rhine rimase affascinato da Lady Wonder, un cavallo sapiente della Virginia di cui descrisse le capacità psi in un entusiastico articolo redatto in collaborazione con la moglie, comparso sul *Journal of Abnormal and Social Psychology* (Vol. 23, N. 4, pagg. 449-466; gennaio-marzo 1929). Successivamente un Rhine più smaliziato ripeté i propri test con Lady Wonder scoprendo i segnali dell'istruttore. Attualmente Rhine sostiene che Lady Wonder disponeva in un primo tempo di autentiche capacità extrasensoriali, e che le segnalazioni segrete cominciarono quando più tardi le perse. (Per i dettagli si veda il mio *Fads and Fallacies in Name of Science*, un paperback della Dover, pagg. 351-352.)

Leggendo i rapporti ufficiali di esperimenti parapsicologici è normalmente impossibile venire a sapere se erano presenti delle spalle. Se prendete gli scritti di Harold E. Puthoff e Russel Targ in cui relazionano sui loro esperimenti di chiaroveggenza con il medium israeliano Uri Geller, all'epoca in cui era ospite dello Stanford Research Institute, non troverete parola del fatto che il migliore amico di Geller, Shipi Strang, era sempre presente ai test. Quando questo fatto venne fuori, essi ammisero che Strang era presente, ma aggiunsero che era «scrupolosamente escluso dall'area di prova». Con ciò intendevano dire che egli non era nella stanza dove venivano scelte casualmente le immagini-bersaglio che Uri doveva «sentire». Ma, attenzione, Strang non era controllato e si poteva muovere liberamente; ci sono molti modi in cui avrebbe potuto identificare i bersagli e segnalarli a Geller. Il contrasto tra l'etereo, asettico rapporto sull'esperimento e le caotiche condizioni in cui si era effettivamente svolto, descritte da John Wilhelms in *The Search for Superman* (Pocket Books, 1976) e da James Randi (*The Magic of Uri Geller*, Ballantine Books, 1976) è sorprendente. Che Strang abbia spesso fatto da spalla a Geller è stato provato, avendolo ammesso sia la sorella di Strang in un'intervista concessa a un reporter israeliano, sia il primo *manager* di Geller, Yasha Katz, in un'intervista trasmessa l'anno scorso alla televisione italiana.

Se un medium deve indovinare delle carte che gli sono tenute nascoste, non è necessario che siano marcate o orientate, perché la spalla gli comunichi informazioni significative. È sufficiente che gli segnali il simbolo dopo che la carta è stata scoperta e il medium otterrà un punteggio superiore a quello casuale. Con delle informazioni complete su ogni carta dopo che è stata «testata», un medium che adotti una strategia ottimale può elevare il suo punteggio atteso da 5 a 8,65 (si veda l'articolo *Cardguessing with Information, a Problem in Probability*, di Ronald C. Read, in «The American Mathematical Monthly», Vol. 69, N. 6, pagg. 506-511, giugno-luglio 1962).

Se il feedback è costituito dalla sola informazione che il medium ha indovinato o ha sbagliato, c'è una strategia che consente di ottenere un punteggio di 6,63 (si veda l'articolo *Statistical Problems in ESP Research*, di Persi Diaconis, in «Science», Vol. 201, N. 4351, pagg. 131-136, luglio 1978). In molti testi classici di ESP non venne fornito alcun ragguaglio circa il fatto che venissero date delle informazioni al medium o che fossero presenti degli amici che avrebbero potuto fornirgli l'opportuno feedback.

Nella maggior parte dei giochi di carte l'informazione fornita dalle carte giocate è di grande valore per un giocatore esper-

Pubblicità per un maiale sapiente in un manifesto della collezione di Milbourn Christopher.

CINQUE DI CUORI — TRE DI PICCHE — DORSO ROSSO — SEI DI PICCHE

Il trucco della predizione della carta rossa.

to. Ciò è chiaro nel caso del bridge e dello stud poker. Nel blackjack, o ventuno, un giocatore in grado di memorizzare il valore delle carte giocate può effettivamente vincere somme consistenti puntando forte quando le probabilità lo favoriscono e poco quando gli sono contro. Questi giocatori sono detti contabili, e negli ultimi anni sono andati così bene che la maggior parte dei casinò si rifiutano di ammettere al gioco i contabili più noti. Ciò non sembra molto corretto perché chi calcola non imbroglia in alcun ragionevole senso della parola.

L'intuizione può smarrirsi notevolmente nella valutazione del feedback fornito dalla distribuzione delle carte. Non conosco un esempio più sorprendente del gioco di carte ideato da Robert Connelly, un matematico della Cornell University, che l'anno scorso fece notizia per la confutazione di una famosa congettura sui poliedri. Se un poliedro ha le facce rigide, ma tutti i suoi angoli hanno delle cerniere, può venire «flesso», ossia può venire modificata la sua forma? Fino dal 1813 è ben noto che la risposta è no se il poliedro è convesso, ed era stato congetturato che lo stesso valesse anche per tutti i poliedri non convessi senza facce che si intersecano. Connelly ha trovato un controesempio con 18 facce triangolari (per una descrizione non tecnica si veda il suo articolo *A Flexible Sphere* in «The Mathematical Intelligencer» Vol. 1, N. 3, pagg. 130-131, 1978).

Connolly chiama il suo gioco Dico Rosso. Il banco mescola un mazzo standard di 52 carte e lentamente le distribuisce a faccia in su. Le carte distribuite sono lasciate in vista dove possono venire controllate dal giocatore in ogni momento. Quando il giocatore vuole dice «Rosso»: se la carta successiva è rossa ha vinto, altrimenti ha perso. Egli deve chiamare il rosso prima della fine della distribuzione, anche se aspetta l'ultima carta per chiamare. A quanto dovrebbe dare la posta il banco per rendere il gioco accettabile, assumendo che il giocatore adotti la strategia migliore sulla base delle informazioni tratte dalle carte giocate? Il giocatore deve dichiarare l'entità della sua scommessa prima dell'inizio di ogni mano. Il prossimo mese fornirò la sorprendente soluzione di Connelly.

Un altro segreto ben custodito dai falsi medium per aumentare la loro probabilità di successo è il ricorso a ciò che Diaconis, nell'articolo sopra citato, chiama punti terminali multipli. I maghi le chiamano più familiarmente «uscite». L'idea di fondo è che il medium non specifica in anticipo quella che sarà la sua strategia, sicché il risultato dipenderà da ciò che succede. Come nota Diaconis le probabilità di una qualche coincidenza sono molto superiori a quelle di una coincidenza specifica. Diaconis descrive l'opera di un medium noto nella letteratura come «B. D.», le cui imprese con mazzi di carte erano favorevolmente commentate in tre saggi del «Journal of Parapsychology» di Rhine. Diaconis, che un tempo era un prestigiatore di grande abilità con le carte, era presente a una delle dimostrazioni di B. D. alla Harvard University. Diaconis si accorse subito che il segreto di B. D. era la sua fiducia nelle uscite.

Anche se non posso descrivere qui nei dettagli il modo in cui i maghi che pretendono di essere medium fanno ampio uso delle uscite, posso però illustrare la tecnica con un esempio tratto da due notevoli trucchi di «precognizione», con cui il lettore potrà stupire gli amici. Paul Curry, un mago dilettante di New York, ama ricordare quella volta che a un party un amico mago era stato pregato di fare qualche gioco con le carte. Questi aveva notato che nel mazzo datogli c'era una carta malamente strappata, diciamo l'otto di fiori. Poiché avrebbe potuto disturbarlo nelle sue manipolazioni, la tolse di nascosto dal mazzo mettendosela in tasca. Nel corso della sua esibizione chiese a qualcuno di nominare una carta. Una donna esclamò: «L'otto di fiori». Senza pensare egli rispose: «Mi spiace, ma dovrebbe chiamare un'altra carta. Prima di iniziare ho notato che l'otto di fiori era strappato e così...». Si interruppe di colpo, dispiaciuto, quando si accorse di aver perso una stupenda occasione di operare un miracolo: far sparire la carta dal mazzo per materializzarla nella tasca, sia pure danneggiandola nel passaggio.

È difficile da credere, ma i maghi hanno elaborato effettivamente dei trucchi con dei mazzi trattati (ordinati in un modo prestabilito) con 52 differenti uscite a seconda della carta scelta. La carta può mostrarsi come quella superiore, quella inferiore o come l'unica girata del mazzo. Uno può scoprirne un doppione attaccato al soffitto, estrarre dal taschino della giacca un fazzoletto con l'immagine della carta, invitare uno spettatore a guardare sotto il suo sedile, dove c'è un doppione, mostrare che è l'unica carta mancante dal mazzo, tirare il mazzo contro le tendine di una finestra per poi aprirle e mostrare la carta attaccata fuori dalla finestra, e così via.

Il seguente ingegnoso trucco, che richiede sei carte e sei uscite, ha diverse varianti. Quella qui esposta è dovuta a Tom Ransom, di Toronto. L'illustrazione in alto indica come sono disposte in fila le carte. Potete informare il vostro uditorio che i valori vanno da 1 a 6. Tutte le carte eccetto una hanno il dorso blu. Tutte le carte sono picche all'infuori del cinque di cuori, che naturalmente è coperto. Scrivete su di un foglio di carta «Sceglierai la carta rossa» e quindi mettete via la predizione.

Chiedete un numero da uno a sei. Se preferite potete porgere a qualcuno un «dado invisibile» e fingere di pregarlo di

tirarlo e di dirvi il risultato. Queste sono le sei uscite per ognuno dei numeri:

1. Contate la prima carta partendo da sinistra; giratela per mostrare che è rossa. Scoprite anche le altre carte coperte per mostrare che sono tutte nere.

2. Girate il due per mostrare il suo dorso rosso e le altre carte scoperte che hanno invece il dorso blu.

3. Pregate qualcuno di contare la terza carta dalla sua sinistra. Quindi come sopra.

4. Contate la quarta carta da sinistra, e quindi come sopra.

5. Scoprite tutte le carte coperte per mostrare che solo il cinque è rosso.

6. Pregate qualcuno di contare da sinistra la sesta carta, quindi terminate come sopra.

Quale che sia il numero scelto, la vostra predizione sarà esatta. Ovviamente non potrete ripetere il trucco con lo stesso pubblico.

Un altro ingegnoso trucco è in vendita nei negozi giapponesi. È invenzione di Shigeo Futigawa, professore di matematica e mago dilettante.

Avete bisogno di quattro carte uguali, bianche da entrambi i lati. Su una di esse scrivete il numero 17 e sul retro il numero 30. Sulle carte rimanenti segnate le seguenti coppie di numeri 26/39, 28/41, 45/58. Nella prima riga dell'illustrazione in basso della pagina a fronte sono disegnate quattro carte che sono state disposte in modo da mostrare il valore più basso. Dovete memorizzare questi quattro numeri in modo da poterli identificare facilmente oppure, se preferite, potete segnare un puntino in modo da distinguerli dai numeri più alti del dorso.

Date le quattro carte a qualcuno pregandolo di mescolarle con cura, girandole come gli piaccia. Si dispongano poi le carte sul tavolo, in modo che i numeri visibili siano scelti casualmente. Prima che ciò venga fatto scrivete la vostra predizione su di un foglio e mettetela da parte.

Quando le carte sono state disposte sul tavolo, si danno tre casi. Per ognuno di essi si segue una procedura differente:

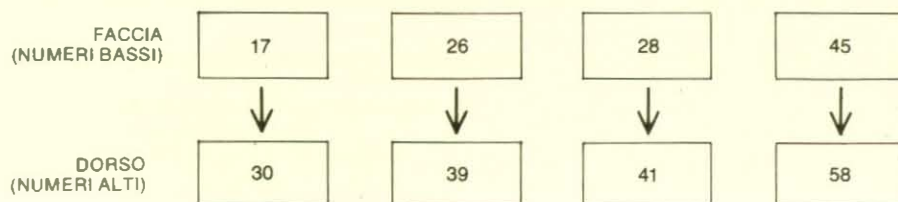
1. Le carte esibiscono due numeri bassi e due alti. In una lunga serie di mani ciò si verificherà tre volte su otto. Sommate i quattro numeri: il risultato sarà 142. Ciò è quanto voi avrete scirtto sul foglio. Pregate qualcuno di controllare.

2. Tutte e quattro le carte sono alte o basse. Ciò avverrà una volta su otto. Alzatevi, girate la schiena e pregate qualcuno di rendere i numeri ancora più casuali girando una coppia di carte. Ciò porta naturalmente a due valori alti e due bassi, di modo che siete nuovamente nella situazione di prima.

3. Le carte mostrano tre valori alti e uno basso oppure tre bassi e uno alto. Questa è peraltro la situazione più comune, che occorre con una probabilità di 1/2. In questo caso prendete la carta solitaria (l'unica bassa o l'unica alta) e girandola dite: «Vi prego di osservare che i numeri sulle facce opposte delle carte sono differenti, sicché, se aveste collocato questa carta rovesciata, ciò avrebbe portato a una somma dei valori visibili differente». Alzatevi, giratevi e chiedete a qualcuno di voltare due carte a piacere per rendere il risultato ancora più casuale. Poiché ciò porta a un'equa divisione fra valori alti e bassi siete ancora nel caso 1.

In breve, la procedura adottata conduce sempre alla somma 142, e non potete sbagliarvi. Osservate che la differenza fra ogni coppia di numeri di una carta è 13. Lascio al lettore la dimostrazione algebrica del trucco. La differenza può essere un numero qualsiasi. La predizione sarà data dalla somma dei quattro valori bassi più il doppio della differenza. In questo modo è facile preparare delle carte con numeri diversi da quelli che ho dato.

È possibile approntare quattro carte con otto valori differenti in modo che il trucco funzioni ancora, con la sola differenza che i numeri scelti vanno moltiplicati e non sommati? Dovete essere in grado di predire il prodotto allo stesso modo della somma, e con le stesse tre uscite. La risposta è sì e Futigawa ha progettato un insieme di carte come variante del trucco precedente. Il mese prossimo lo descriverò, spiegandone il segreto.

Due osservazioni su quanto ho scritto circa i cerchi tangenti nel numero di maggio. Benjamin L. Schwartz mi comunica che una dimostrazione della disposizione ottimale di sei cerchi in un quadrato è apparsa per la prima volta nel suo articolo *Separating Points in a Square*, sul «Journal of Recreational Mathematics» (Vol. 3, pagg. 195-204, ottobre 1970). Il problema della sistemazione dei cerchi, così come l'ho formulato, è equivalente a quello di sistemare n punti all'interno di un quadrato in modo che la distanza minima fra ogni coppia di punti sia massimizzata. La mia descrizione della spirale di cerchi tangenti di H. S. M. Coxeter come «unica» era erronea. Infatti essa è unica solo quando la ragione di due raggi consecutivi è quella specificata. Per i dettagli si veda lo scritto di Coxeter citato nella bibliografia di questo numero.

Il trucco della predizione della somma di Shigeo Futigawa.